

**SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE
SAME**

FIELD OF THE INVENTION

The present invention relates generally to the identification and isolation of novel DNA and to the
5 recombinant production of novel polypeptides.

BACKGROUND OF THE INVENTION

Extracellular proteins play important roles in, among other things, the formation, differentiation and maintenance of multicellular organisms. The fate of many individual cells, e.g., proliferation, migration, differentiation, or interaction with other cells, is typically governed by information received from other cells and/or the immediate environment. This information is often transmitted by secreted polypeptides (for instance, mitogenic factors, survival factors, cytotoxic factors, differentiation factors, neuropeptides, and hormones) which are, in turn, received and interpreted by diverse cell receptors or membrane-bound proteins. These secreted polypeptides or signaling molecules normally pass through the cellular secretory pathway to reach their site of action in the extracellular environment.

Secreted proteins have various industrial applications, including as pharmaceuticals, diagnostics, biosensors and bioreactors. Most protein drugs available at present, such as thrombolytic agents, interferons, interleukins, erythropoietins, colony stimulating factors, and various other cytokines, are secretory proteins. Their receptors, which are membrane proteins, also have potential as therapeutic or diagnostic agents. Efforts are being undertaken by both industry and academia to identify new, native secreted proteins. Many efforts are focused on the screening of mammalian recombinant DNA libraries to identify the coding sequences for novel secreted proteins. Examples of screening methods and techniques are described in the literature [see, for example, Klein et al., *Proc. Natl. Acad. Sci.* 93:7108-7113 (1996); U.S. Patent No. 5,536,637].

Membrane-bound proteins and receptors can play important roles in, among other things, the formation, differentiation and maintenance of multicellular organisms. The fate of many individual cells, e.g., proliferation, migration, differentiation, or interaction with other cells, is typically governed by information received from other cells and/or the immediate environment. This information is often transmitted by secreted polypeptides (for instance, mitogenic factors, survival factors, cytotoxic factors, differentiation factors, neuropeptides, and hormones) which are, in turn, received and interpreted by diverse cell receptors or membrane-bound proteins. Such membrane-bound proteins and cell receptors include, but are not limited to, cytokine receptors, receptor kinases, receptor phosphatases, receptors involved in cell-cell interactions, and cellular adhesin molecules like selectins and integrins. For instance, transduction of signals that regulate cell growth and differentiation is regulated in part by phosphorylation of various cellular proteins. Protein tyrosine kinases, enzymes that catalyze

that process, can also act as growth factor receptors. Examples include fibroblast growth factor receptor and nerve growth factor receptor.

Membrane-bound proteins and receptor molecules have various industrial applications, including as pharmaceutical and diagnostic agents. Receptor immunoadhesins, for instance, can be employed as therapeutic agents to block receptor-ligand interactions. The membrane-bound proteins can also be employed for screening 5 of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction.

Efforts are being undertaken by both industry and academia to identify new, native receptor or membrane-bound proteins. Many efforts are focused on the screening of mammalian recombinant DNA libraries to identify the coding sequences for novel receptor or membrane-bound proteins.

10 1. **PRO241**

Cartilage is a specialized connective tissue with a large extracellular matrix containing a dense network of collagen fibers and a high content of proteoglycan. While the majority of the proteoglycan in cartilage is aggrecan, which contains many chondroitin sulphate and keratin sulphate chains and forms multimolecular aggregates by binding with link protein to hyaluronan, cartilage also contains a number of smaller molecular weight proteoglycans. One of these smaller molecular weight proteoglycans is a protein called biglycan, a proteoglycan which is widely distributed in the extracellular matrix of various other connective tissues including tendon, sclera, skin, and the like. Biglycan is known to possess leucine-rich repeat sequences and two chondroitin sulphate/dermatan sulphate chains and functions to bind to the cell-binding domain of fibronectin so as to inhibit cellular attachment thereto. It is speculated that the small molecular weight proteoglycans such as biglycan may play important roles in the growth and/or repair of cartilage and in degenerative diseases such as arthritis. As such, there is an interest in identifying and characterizing novel polypeptides having homology to biglycan protein.

We herein describe the identification and characterization of novel polypeptides having homology to the biglycan protein, wherein those polypeptides are herein designated PRO241 polypeptides.

25

2. **PRO243**

Chordin (*Xenopus*, Xchd) is a soluble factor secreted by the Spemann organizer which has potent dorsalizing activity (Sasai *et al.*, *Cell* 79: 779-90 (1994); Sasai *et al.*, *Nature* 376: 333-36 (1995). Other dorsalizing factors secreted by the organizer are noggin (Smith and Harlan, *Cell* 70: 829-840 (1992); Lamb *et al.*, *Science* 262: 713-718 (1993) and follistatin (Hemmant-Briantou *et al.*, *Cell* 77: 283-295 (1994). Chordin 30 subdivides primitive ectoderm into neural versus non-neural domains, and induces notochord and muscle formation by the dorsalization of the mesoderm. It does this by functioning as an antagonist of the ventralizing BMP-4 signals. This inhibition is mediated by direct binding of chordin to BMP-4 in the extracellular space, thereby preventing BMP-4 receptor activation by BMP-4 (Piccolo *et al.*, *Develop. Biol.* 182: 5-20 (1996).

35 BMP-4 is expressed in a gradient from the ventral side of the embryo, while chordin is expressed in a gradient complementary to that of BMP-4. Chordin antagonizes BMP-4 to establish the low end of the BMP gradient. Thus, the balance between the signal from chordin and other organizer-derived factors versus the BMP

signal provides the ectodermal germ layer with its dorsal-ventral positional information. Chordin may also be involved in the dorsal-ventral patterning of the central nervous system (Sasai *et al*, *Cell* 79: 779-90 (1994). It also induces exclusively anterior neural tissues (forebrain-type), thereby anteriorizing the neural type (Sasai *et al*, *Cell* 79: 779-90 (1997). Given its role in neuronal induction and patterning, chordin may prove useful in the treatment of neurodegenerative disorders and neural damage, e.g., due to trauma or after chemotherapy.

5 We herein describe the identification and characterization of novel polypeptides having homology to the chordin protein, wherein those polypeptides are herein designated PRO243 polypeptides.

3. **PRO299**

The notch proteins are involved in signaling during development. They may effect asymmetric development potential and may signal expression of other proteins involved in development. [See Robey, E., *Curr. Opin. Genet. Dev.*, 7(4):551 (1997), Simpson, P., *Curr. Opin. Genet. Dev.*, 7(4):537 (1997), Blobel, CP., *Cell*, 90(4):589 (1997)], Nakayama, H. et al., *Dev. Genet.*, 21(1):21 (1997), Nakayama, H. et al., *Dev. Genet.*, 21(1):21 (1997), Sullivan, S.A. et al., *Dev. Genet.*, 20(3):208 (1997) and Hayashi, H. et al., *Int. J. Dev. Biol.*, 40(6):1089(1996).] Serrate-mediated activation of notch has been observed in the dorsal compartment of the Drosophila wing imaginal disc. Fleming et al., *Development*, 124(15):2973 (1997). Notch is of interest for both its role in development as well as its signaling abilities. Also of interest are novel polypeptides which may have a role in development and/or signaling.

We herein describe the identification and characterization of novel polypeptides having homology to the notch protein, wherein those polypeptides are herein designated PRO299 polypeptides.

4. **PRO323**

Dipeptidases are enzymatic proteins which function to cleave a large variety of different dipeptides and which are involved in an enormous number of very important biological processes in mammalian and non-mammalian organisms. Numerous different dipeptidase enzymes from a variety of different mammalian and non-mammalian organisms have been both identified and characterized. The mammalian dipeptidase enzymes play important roles in many different biological processes including, for example, protein digestion, activation, inactivation, or modulation of dipeptide hormone activity, and alteration of the physical properties of proteins and enzymes.

In light of the important physiological roles played by dipeptidase enzymes, efforts are being undertaken by both industry and academia to identify new, native dipeptidase homologs. Many efforts are focused on the screening of mammalian recombinant DNA libraries to identify the coding sequences for novel secreted and membrane-bound receptor proteins. Examples of screening methods and techniques are described in the literature [see, for example, Klein et al., *Proc. Natl. Acad. Sci.*, 93:7108-7113 (1996); U.S. Patent No. 5,536,637].

35 We herein describe the identification and characterization of novel polypeptides having homology to various dipeptidase enzymes, designated herein as PRO323 polypeptides.

5. **PRO327**

The anterior pituitary hormone prolactin is encoded by a member of the growth hormone/prolactin/placental lactogen gene family. In mammals, prolactin is primarily responsible for the development of the mammary gland and lactation. Prolactin functions to stimulate the expression of milk protein genes by increasing both gene transcription and mRNA half-life.

5 The physiological effects of the prolactin protein are mediated through the ability of prolactin to bind to a cell surface prolactin receptor. The prolactin receptor is found in a variety of different cell types, has a molecular mass of approximately 40,000 and is apparently not linked by disulfide bonds to itself or to other subunits. Prolactin receptor levels are differentially regulated depending upon the tissue studied.

Given the important physiological roles played by cell surface receptor molecules *in vivo*, efforts are 10 currently being undertaken by both industry and academia to identify new, native membrane-bound receptor proteins, including those which share sequence homology with the prolactin receptor. Many of these efforts are focused on the screening of mammalian recombinant DNA libraries to identify the coding sequences for novel membrane-bound receptor proteins. Examples of screening methods and techniques are described in the literature [see, for example, Klein et al., *Proc. Natl. Acad. Sci.*, 93:7108-7113 (1996); U.S. Patent No. 15 5,536,637].

We herein describe the identification and characterization of novel polypeptides having significant homology to the prolactin receptor protein, designated herein as PRO327 polypeptides.

6. **PRO233**

Studies have reported that the redox state of the cell is an important determinant of the fate of the cell. Furthermore, reactive oxygen species have been reported to be cytotoxic, causing inflammatory disease, including tissue necrosis, organ failure, atherosclerosis, infertility, birth defects, premature aging, mutations and malignancy. Thus, the control of oxidation and reduction is important for a number of reasons, including the control and prevention of strokes, heart attacks, oxidative stress and hypertension.

25 Oxygen free radicals and antioxidants appear to play an important role in the central nervous system after cerebral ischemia and reperfusion. Moreover, cardiac injury, related to ischaemia and reperfusion has been reported to be caused by the action of free radicals. In this regard, reductases, and particularly, oxidoreductases, are of interest. In addition, the transcription factors, NF-kappa B and AP-1, are known to be regulated by redox state and to affect the expression of a large variety of genes thought to be involved in the pathogenesis of AIDS, 30 cancer, atherosclerosis and diabetic complications. Publications further describing this subject matter include Kelsey et al., *Br. J. Cancer*, 76(7):852-854 (1997); Friedrich and Weiss, *J. Theor. Biol.*, 187(4):529-540 (1997) and Pieulle et al., *J. Bacteriol.*, 179(18):5684-5692 (1997). Given the physiological importance of redox reactions *in vivo*, efforts are currently being undertaken to identify new, native proteins which are involved in redox reactions. We describe herein the identification and characterization of novel polypeptides which have 35 homology to reductase, designated herein as PRO233 polypeptides.

7. **PRO344**

The complement proteins comprise a large group of serum proteins some of which act in an enzymatic cascade, producing effector molecules involved in inflammation. The complement proteins are of particular physiological importance in regulating movement and function of cells involved in inflammation. Given the physiological importance of inflammation and related mechanisms *in vivo*, efforts are currently being undertaken

5 to identify new, native proteins which are involved in inflammation. We describe herein the identification and characterization of novel polypeptides which have homology to complement proteins, wherein those polypeptides are herein designated as PRO344 polypeptides.

8. **PRO347**

10 Cysteine-rich proteins are generally proteins which have intricate three-dimensional structures and/or exist in multimeric forms due to the presence of numerous cysteine residues which are capable of forming disulfide bridges. One well known cysteine-rich protein is the mannose receptor which is expressed in, among other tissues, liver where it serves to bind to mannose and transport it into liver cells. Other cysteine-rich proteins are known to play important roles in many other physiological and biochemical processes. As such, there is an interest in identifying novel cysteine-rich proteins. In this regard, Applicants describe herein the identification and characterization of novel cysteine-rich polypeptides that have significant sequence homology to the cysteine-rich secretory protein-3, designated herein as PRO347 polypeptides.

15
20
25
30
35

9. **PRO354**

20 Inter-alpha-trypsin inhibitor (ITI) is a large (Mr approximately 240,000) circulating protease inhibitor found in the plasma of many mammalian species. The intact inhibitor is a glycoprotein and consists of three glycosylated subunits that interact through a strong glycosaminoglycan linkage. The anti-trypsin activity of ITI is located on the smallest subunit (i.e., the light chain) of the complex, wherein that light chain is now known as the protein bikunin. The mature light chain consists of a 21-amino acid N-terminal sequence, glycosylated at Ser-10, followed by two tandem Kunitz-type domains, the first of which is glycosylated at Asn-45 and the second of which is capable of inhibiting trypsin, chymotrypsin and plasmin. The remaining two chains of the ITI complex are heavy chains which function to interact with the enzymatically active light chain of the complex.

25 Efforts are being undertaken by both industry and academia to identify new, native proteins. Many efforts are focused on the screening of mammalian recombinant DNA libraries to identify the coding sequences for novel secreted and membrane-bound receptor proteins. Examples of screening methods and techniques are described in the literature [see, for example, Klein et al., *Proc. Natl. Acad. Sci.*, 93:7108-7113 (1996); U.S. Patent No. 5,536,637]. We herein describe the identification and characterization of novel polypeptides having significant homology to the ITI heavy chain, designated in the present application as PRO354 polypeptides.

35 10. **PRO355**

Cytotoxic or regulatory T cell associated molecule or "CRTAM" protein is structurally related to the immunoglobulin superfamily. The CRTAM protein should be capable of mediating various immune responses.

Antibodies typically bind to CRTAM proteins with high affinity. Zlotnik, A., Faseb, 10(6): A1037, Abr. 216, June 1996. Given the physiological importance of T cell antigens and immune processes *in vivo*, efforts are currently being undertaken to identify new, native proteins which are involved in immune responses. See also Kennedy et al., U.S. Pat. No. 5,686,257 (1997). We describe herein the identification and characterization of novel polypeptides which have homology to CRTAM, designated in the present application as PRO355

5 polypeptides.

11. **PRO357**

Protein-protein interactions include receptor and antigen complexes and signaling mechanisms. As more is known about the structural and functional mechanisms underlying protein-protein interactions, protein-protein 10 interactions can be more easily manipulated to regulate the particular result of the protein-protein interaction. Thus, the underlying mechanisms of protein-protein interactions are of interest to the scientific and medical community.

All proteins containing leucine-rich repeats are thought to be involved in protein-protein interactions. Leucine-rich repeats are short sequence motifs present in a number of proteins with diverse functions and cellular 15 locations. The crystal structure of ribonuclease inhibitor protein has revealed that leucine-rich repeats correspond to beta-alpha structural units. These units are arranged so that they form a parallel beta-sheet with one surface exposed to solvent, so that the protein acquires an unusual, nonglobular shape. These two features have been indicated as responsible for the protein-binding functions of proteins containing leucine-rich repeats. See, Kobe and Deisenhofer, Trends Biochem. Sci., 19(10):415-421 (Oct. 1994).

A study has been reported on leucine-rich proteoglycans which serve as tissue organizers, orienting and ordering collagen fibrils during ontogeny and are involved in pathological processes such as wound healing, tissue repair, and tumor stroma formation. Iozzo, R. V., Crit. Rev. Biochem. Mol. Biol., 32(2):141-174 20 (1997). Others studies implicating leucine rich proteins in wound healing and tissue repair are De La Salle, C., et al., Vouv. Rev. Fr. Hematol. (Germany), 37(4):215-222 (1995), reporting mutations in the leucine rich motif in a complex associated with the bleeding disorder Bernard-Soulier syndrome, Chlemetson, K. J., Thromb. 25 Haemost. (Germany), 74(1):111-116 (July 1995), reporting that platelets have leucine rich repeats and Ruoslahti, E. I., et al., WO9110727-A by La Jolla Cancer Research Foundation reporting that decorin binding to transforming growth factor β has involvement in a treatment for cancer, wound healing and scarring. Related by function to this group of proteins is the insulin like growth factor (IGF), in that it is useful in wound-healing and 30 associated therapies concerned with re-growth of tissue, such as connective tissue, skin and bone; in promoting body growth in humans and animals; and in stimulating other growth-related processes. The acid labile subunit (ALS) of IGF is also of interest in that it increases the half-life of IGF and is part of the IGF complex in vivo.

Another protein which has been reported to have leucine-rich repeats is the SLIT protein which has been 35 reported to be useful in treating neuro-degenerative diseases such as Alzheimer's disease, nerve damage such as in Parkinson's disease, and for diagnosis of cancer, see, Artavanistsakonas, S. and Rothberg, J. M., WO9210518-A1 by Yale University. Also of interest is LIG-1, a membrane glycoprotein that is expressed specifically in glial cells in the mouse brain, and has leucine rich repeats and immunoglobulin-like domains.

Suzuki, et al., *J. Biol. Chem. (U.S.)*, 271(37):22522 (1996). Other studies reporting on the biological functions of proteins having leucine rich repeats include: Tayar, N., et al., *Mol. Cell Endocrinol.*, (Ireland), 125(1-2):65-70 (Dec. 1996) (gonadotropin receptor involvement); Miura, Y., et al., *Nippon Rinsho* (Japan), 54(7):1784-1789 (July 1996) (apoptosis involvement); Harris, P. C., et al., *J. Am. Soc. Nephrol.*, 6(4):1125-1133 (Oct. 1995) (kidney disease involvement).

5 Efforts are therefore being undertaken by both industry and academia to identify new proteins having leucine rich repeats to better understand protein-protein interactions. Of particular interest are those proteins having leucine rich repeats and homology to known proteins having leucine rich repeats such as the acid labile subunit of insulin-like growth factor. Many efforts are focused on the screening of mammalian recombinant DNA libraries to identify the coding sequences for novel secreted and membrane-bound proteins having leucine 10 rich repeats. Examples of screening methods and techniques are described in the literature [see, for example, Klein et al., *Proc. Natl. Acad. Sci.*, 93:7108-7113 (1996); U.S. Patent No. 5,536,637].

We describe herein the identification and characterization of novel polypeptides having homology to the acid labile subunit of insulin-like growth factor, designated in the present application as PRO357 polypeptides.

15 12. **PRO715**

Control of cell numbers in mammals is believed to be determined, in part, by a balance between cell proliferation and cell death. One form of cell death, sometimes referred to as necrotic cell death, is typically characterized as a pathologic form of cell death resulting from some trauma or cellular injury. In contrast, there is another, "physiologic" form of cell death which usually proceeds in an orderly or controlled manner. This orderly or controlled form of cell death is often referred to as "apoptosis" [see, e.g., Barr et al., *Bio/Technology*, 12:487-493 (1994); Steller et al., *Science*, 267:1445-1449 (1995)]. Apoptotic cell death naturally occurs in many physiological processes, including embryonic development and clonal selection in the immune system [Itoh et al., *Cell*, 66:233-243 (1991)]. Decreased levels of apoptotic cell death have been associated with a variety of pathological conditions, including cancer, lupus, and herpes virus infection 25 [Thompson, *Science*, 267:1456-1462 (1995)]. Increased levels of apoptotic cell death may be associated with a variety of other pathological conditions, including AIDS, Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, multiple sclerosis, retinitis pigmentosa, cerebellar degeneration, aplastic anemia, myocardial infarction, stroke, reperfusion injury, and toxin-induced liver disease [see, Thompson, *supral.*].

Apoptotic cell death is typically accompanied by one or more characteristic morphological and 30 biochemical changes in cells, such as condensation of cytoplasm, loss of plasma membrane microvilli, segmentation of the nucleus, degradation of chromosomal DNA or loss of mitochondrial function. A variety of extrinsic and intrinsic signals are believed to trigger or induce such morphological and biochemical cellular changes [Raff, *Nature*, 356:397-400 (1992); Steller, *supra*; Sachs et al., *Blood*, 82:15 (1993)]. For instance, they can be triggered by hormonal stimuli, such as glucocorticoid hormones for immature thymocytes, as well 35 as withdrawal of certain growth factors [Watanabe-Fukunaga et al., *Nature*, 356:314-317 (1992)]. Also, some identified oncogenes such as *myc*, *rel*, and *EIA*, and tumor suppressors, like *p53*, have been reported to have a role in inducing apoptosis. Certain chemotherapy drugs and some forms of radiation have likewise been

observed to have apoptosis-inducing activity [Thompson, supra].

Various molecules, such as tumor necrosis factor- α ("TNF- α "), tumor necrosis factor- β ("TNF- β " or "lymphotoxin- α "), lymphotoxin- β ("LT- β "), CD30 ligand, CD27 ligand, CD40 ligand, OX-40 ligand, 4-1BB ligand, Apo-1 ligand (also referred to as Fas ligand or CD95 ligand), and Apo-2 ligand (also referred to as TRAIL) have been identified as members of the tumor necrosis factor ("TNF") family of cytokines [See, e.g., 5 Gruss and Dower, Blood, 85:3378-3404 (1995); Pitti et al., J. Biol. Chem., 271:12687-12690 (1996); Wiley et al., Immunity, 3:673-682 (1995); Browning et al., Cell, 72:847-856 (1993); Armitage et al. Nature, 357:80-82 (1992)]. Among these molecules, TNF- α , TNF- β , CD30 ligand, 4-1BB ligand, Apo-1 ligand, and Apo-2 ligand (TRAIL) have been reported to be involved in apoptotic cell death. Both TNF- α and TNF- β have been reported to induce apoptotic death in susceptible tumor cells [Schmid et al., Proc. Natl. Acad. Sci., 83:1881 10 (1986); Dealtry et al., Eur. J. Immunol., 17:689 (1987)]. Zheng et al. have reported that TNF- α is involved in post-stimulation apoptosis of CD8-positive T cells [Zheng et al., Nature, 377:348-351 (1995)]. Other investigators have reported that CD30 ligand may be involved in deletion of self-reactive T cells in the thymus [Amakawa et al., Cold Spring Harbor Laboratory Symposium on Programmed Cell Death, Abstr. No. 10, 15 (1995)].

20 Mutations in the mouse Fas/Apo-1 receptor or ligand genes (called *lpr* and *gld*, respectively) have been associated with some autoimmune disorders, indicating that Apo-1 ligand may play a role in regulating the clonal deletion of self-reactive lymphocytes in the periphery [Krammer et al., Curr. Op. Immunol., 6:279-289 (1994); Nagata et al., Science, 267:1449-1456 (1995)]. Apo-1 ligand is also reported to induce post-stimulation apoptosis in CD4-positive T lymphocytes and in B lymphocytes, and may be involved in the elimination of activated lymphocytes when their function is no longer needed [Krammer et al., supra; Nagata et al., supra]. 25 Agonist mouse monoclonal antibodies specifically binding to the Apo-1 receptor have been reported to exhibit cell killing activity that is comparable to or similar to that of TNF- α [Yonehara et al., J. Exp. Med., 169:1747-1756 (1989)].

30 Induction of various cellular responses mediated by such TNF family cytokines is believed to be initiated by their binding to specific cell receptors. Two distinct TNF receptors of approximately 55-kDa (TNFR1) and 75-kDa (TNFR2) have been identified [Hohman et al., J. Biol. Chem., 264:14927-14934 (1989); Brockhaus et al., Proc. Natl. Acad. Sci., 87:3127-3131 (1990); EP 417,563, published March 20, 1991] and human and mouse cDNAs corresponding to both receptor types have been isolated and characterized [Loetscher et al., Cell, 61:351 (1990); Schall et al., Cell, 61:361 (1990); Smith et al., Science, 248:1019-1023 (1990); Lewis et al., Proc. Natl. Acad. Sci., 88:2830-2834 (1991); Goodwin et al., Mol. Cell. Biol., 11:3020-3026 (1991)]. The TNF family ligands identified to date, with the exception of lymphotoxin- α , are type II transmembrane proteins, whose C-terminus is extracellular. In contrast, most receptors in the TNF receptor (TNFR) family identified to date are type I transmembrane proteins. In both the TNF ligand and receptor families, however, homology identified between family members has been found mainly in the extracellular domain ("ECD"). Several of the 35 TNF family cytokines, including TNF- α , Apo-1 ligand and CD40 ligand, are cleaved proteolytically at the cell surface; the resulting protein in each case typically forms a homotrimeric molecule that functions as a soluble cytokine. TNF receptor family proteins are also usually cleaved proteolytically to release soluble receptor ECDs

that can function as inhibitors of the cognate cytokines.

Recently, other members of the TNFR family have been identified. Such newly identified members of the TNFR family include CAR1, HVEM and osteoprotegerin (OPG) [Brojatsch et al., Cell, 87:845-855 (1996); Montgomery et al., Cell, 87:427-436 (1996); Marsters et al., J. Biol. Chem., 272:14029-14032 (1997); Simonet et al., Cell, 89:309-319 (1997)]. Unlike other known TNFR-like molecules, Simonet et al., supra, report that

5 OPG contains no hydrophobic transmembrane-spanning sequence.

For a review of the TNF family of cytokines and their receptors, see Gruss and Dower, supra.

Applicants herein describe the identification and characterization of novel polypeptides having homology to members of the tumor necrosis factor family of polypeptides, designated herein as PRO715 polypeptides.

10 13. **PRO353**

The complement proteins comprise a large group of serum proteins some of which act in an enzymatic cascade, producing effector molecules involved in inflammation. The complement proteins are of particular importance in regulating movement and function of cells involved in inflammation. Given the physiological importance of inflammation and related mechanisms *in vivo*, efforts are currently being undertaken to identify new, native proteins which are involved in inflammation. We describe herein the identification and characterization of novel polypeptides which have homology to complement proteins, designated herein as PRO353 polypeptides.

15 14. **PRO361**

The mucins comprise a family of glycoproteins which have been implicated in carcinogenesis. Mucin and mucin-like proteins are secreted by both normal and transformed cells. Both qualitative and quantitative changes in mucins have been implicated in various types of cancer. Given the medical importance of cancer, efforts are currently being undertaken to identify new, native proteins which may be useful for the diagnosis or treatment of cancer.

20 The chitinase proteins comprise a family of which have been implicated in pathogenesis responses in plants. Chitinase proteins are produced by plants and microorganisms and may play a role in the defense of plants to injury. Given the importance of plant defense mechanisms, efforts are currently being undertaken to identify new, native proteins which may be useful for modulation of pathogenesis-related responses in plants. We describe herein the identification and characterization of novel polypeptides which have homology to mucin 30 and chitinase, designated in the present application as PRO361 polypeptides.

25 15. **PRO365**

30 Polypeptides such as human 2-19 protein may function as cytokines. Cytokines are low molecular weight proteins which function to stimulate or inhibit the differentiation, proliferation or function of immune cells. Cytokines often act as intercellular messengers and have multiple physiological effects. Given the physiological importance of immune mechanisms *in vivo*, efforts are currently being undertaken to identify new, native proteins which are involved in effecting the immune system. We describe herein the identification and

characterization of novel polypeptides which have homology to the human 2-19 protein, designated heein as PRO365 polypeptides.

SUMMARY OF THE INVENTION

1. PRO241

5 Applicants have identified a cDNA clone that encodes a novel polypeptide having homology to biglycan protein, wherein the polypeptide is designated in the present application as "PRO241".

In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding a PRO241 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO241 polypeptide having amino acid residues 1 to 379 of Figure 2 (SEQ ID NO:2), or is complementary to such 10 encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions.

In another embodiment, the invention provides isolated PRO241 polypeptide. In particular, the invention provides isolated native sequence PRO241 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 to 379 of Figure 2 (SEQ ID NO:2). Another embodiment of the present invention is directed to a PRO241 polypeptide lacking the N-terminal signal peptide, wherein the PRO241 polypeptide comprises about amino acids 16 to 379 of the full-length PRO241 amino acid sequence (SEQ ID NO:2).

2. PRO243

20 Applicants have identified a cDNA clone (DNA35917-1207) that encodes a novel polypeptide, designated in the present application as "PRO243".

In one embodiment, the invention provides an isolated nucleic acid molecule having at least about 80% sequence identity to (a) a DNA molecule encoding a PRO243 polypeptide comprising the sequence of amino acids 1 or about 24 to 954 of Figure 4 (SEQ ID NO:7), or (b) the complement of the DNA molecule of (a). The 25 sequence identity preferably is about 85%, more preferably about 90%, most preferably about 95%. In one aspect, the isolated nucleic acid has at least about 80%, preferably at least about 85%, more preferably at least about 90%, and most preferably at least about 95% sequence identity with a polypeptide having amino acid residues 1 or about 24 to 954 of Figure 4 (SEQ ID NO:7). Preferably, the highest degree of sequence identity occurs within the four (4) conserved cysteine clusters (amino acids 51 to 125; amino acids 705 to 761; amino 30 acids 784 to 849; and amino acids 897 to 931) of Figure 4 (SEQ ID NO:7). In a further embodiment, the isolated nucleic acid molecule comprises DNA encoding a PRO243 polypeptide having amino acid residues 1 or about 24 to 954 of Figure 4 (SEQ ID NO:7), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions. In another aspect, the invention provides a nucleic acid of the full length protein of clone DNA35917-1207, 35 deposited with the ATCC under accession number ATCC 209508, alternatively the coding sequence of clone DNA35917-1207, deposited under accession number ATCC 209508.

In yet another embodiment, the invention provides isolated PRO243 polypeptide. In particular, the invention provides isolated native sequence PRO243 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 or about 24 to 954 of Figure 4 (SEQ ID NO:7). Native PRO243 polypeptides with or without the native signal sequence (amino acids 1 to 23 in Figure 4 (SEQ ID NO:7)), and with or without the initiating methionine are specifically included. Alternatively, the invention provides a 5 PRO243 polypeptide encoded by the nucleic acid deposited under accession number ATCC 209508.

3. **PRO299**

Applicants have identified a cDNA clone that encodes a novel polypeptide, wherein the polypeptide is designated in the present application as "PRO299".

10 In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding a PRO299 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO299 polypeptide having amino acid residues 1 to 737 of Figure 6 (SEQ ID NO:15), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions.

15 In another embodiment, the invention provides isolated PRO299 polypeptide. In particular, the invention provides isolated native sequence PRO299 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 to 737 of Figure 6 (SEQ ID NO:15). An additional embodiment of the present invention is directed to an isolated extracellular domain of a PRO299 polypeptide.

20 4. **PRO323**

Applicants have identified a cDNA clone that encodes a novel polypeptide having homology to a microsomal dipeptidase protein, wherein the polypeptide is designated in the present application as "PRO323".

25 In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding a PRO323 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO323 polypeptide having amino acid residues 1 to 433 of Figure 10 (SEQ ID NO:24), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions.

30 In another embodiment, the invention provides isolated PRO323 polypeptide. In particular, the invention provides isolated native sequence PRO323 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 to 433 of Figure 10 (SEQ ID NO:24).

5. **PRO327**

Applicants have identified a cDNA clone that encodes a novel polypeptide having homology to prolactin receptor, wherein the polypeptide is designated in the present application as "PRO327".

35 In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding a PRO327 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO327 polypeptide having amino acid residues 1 to 422 of Figure 14 (SEQ ID NO:32), or is complementary to such

encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions.

In another embodiment, the invention provides isolated PRO327 polypeptide. In particular, the invention provides isolated native sequence PRO327 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 to 422 of Figure 14 (SEQ ID NO:32).

5

6. **PRO233**

Applicants have identified a cDNA clone that encodes a novel polypeptide, wherein the polypeptide is designated in the present application as "PRO233".

In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding 10 a PRO233 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO233 polypeptide having amino acid residues 1 to 300 of Figure 16 (SEQ ID NO:37), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions.

In another embodiment, the invention provides isolated PRO233 polypeptide. In particular, the invention provides isolated native sequence PRO233 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 to 300 of Figure 16 (SEQ ID NO:37).

15
16
17
18
19
20
21
22
23
24
25

7. **PRO344**

Applicants have identified a cDNA clone that encodes a novel polypeptide, wherein the polypeptides are designated in the present application as "PRO344".

In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding 20 a PRO344 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO344 polypeptide having amino acid residues 1 to 243 of Figure 18 (SEQ ID NO:42), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions.

In another embodiment, the invention provides isolated PRO344 polypeptide. In particular, the invention provides isolated native sequence PRO344 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 to 243 of Figure 18 (SEQ ID NO:42).

30 8. **PRO347**

Applicants have identified a cDNA clone that encodes a novel polypeptide having homology to cysteine-rich secretory protein-3, wherein the polypeptide is designated in the present application as "PRO347".

In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding 35 a PRO347 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO347 polypeptide having amino acid residues 1 to 455 of Figure 20 (SEQ ID NO:50), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions.

In another embodiment, the invention provides isolated PRO347 polypeptide. In particular, the invention provides isolated native sequence PRO347 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 to 455 of Figure 20 (SEQ ID NO:50).

9. **PRO354**

5 Applicants have identified a cDNA clone that encodes a novel polypeptide having homology to the heavy chain of the inter-alpha-trypsin inhibitor (ITI), wherein the polypeptide is designated in the present application as "PRO354".

In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding a PRO354 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO354 polypeptide having amino acid residues 1 to 694 of Figure 22 (SEQ ID NO:55), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions.

In another embodiment, the invention provides isolated PRO354 polypeptide. In particular, the invention provides isolated native sequence PRO354 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 to 694 of Figure 22 (SEQ ID NO:55).

10. **PRO355**

15 Applicants have identified a cDNA clone that encodes a novel polypeptide, wherein the polypeptide is designated in the present application as "PRO355".

20 In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding a PRO355 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO355 polypeptide having amino acid residues 1 to 440 of Figure 24 (SEQ ID NO:61), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions.

25 In another embodiment, the invention provides isolated PRO355 polypeptide. In particular, the invention provides isolated native sequence PRO355 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 to 440 of Figure 24 (SEQ ID NO:61). An additional embodiment of the present invention is directed to an isolated extracellular domain of a PRO355 polypeptide.

30 11. **PRO357**

Applicants have identified a cDNA clone that encodes a novel polypeptide having homology to insulin-like growth factor (IGF) acid labile subunit (ALS), wherein the polypeptide is designated in the present application as "PRO357".

35 In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding a PRO357 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO357 polypeptide having amino acid residues 1 through 598 of Figure 26 (SEQ ID NO:69), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally,

under high stringency conditions.

In another embodiment, the invention provides isolated PRO357 polypeptide. In particular, the invention provides isolated native sequence PRO357 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 through 598 of Figure 26 (SEQ ID NO:69). An additional embodiment of the present invention is directed to an isolated extracellular domain of a PRO357 polypeptide.

5

12. PRO715

Applicants have identified cDNA clones that encode novel polypeptides having homology to tumor necrosis factor family polypeptides, wherein the polypeptides are designated in the present application as "PRO715".

10 In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding a PRO715 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO715 polypeptide having amino acid residues 1 to 250 of Figure 28 (SEQ ID NO:76), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions.

15 In another embodiment, the invention provides isolated PRO715 polypeptide. In particular, the invention provides isolated native sequence PRO715 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 to 250 of Figure 28 (SEQ ID NO:76). An additional embodiment of the present invention is directed to an isolated extracellular domain of a PRO715 polypeptide.

PRO715
PRO353
PRO361

20 **13. PRO353**

Applicants have identified a cDNA clone that encodes a novel polypeptide, wherein the polypeptides are designated in the present application as "PRO353".

25 In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding a PRO353 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO353 polypeptide having amino acid residues 1 to 281 of Figure 30 (SEQ ID NO:78), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions.

30 In another embodiment, the invention provides an isolated PRO353 polypeptide. In particular, the invention provides isolated native sequence PRO353 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 to 281 of Figure 30 (SEQ ID NO:78).

14. PRO361

Applicants have identified a cDNA clone that encodes a novel polypeptide, wherein the polypeptide is designated in the present application as "PRO361".

35 In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding a PRO361 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO361 polypeptide having amino acid residues 1 to 431 of Figure 32 (SEQ ID NO:83), or is complementary to such

encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions. The isolated nucleic acid sequence may comprise the cDNA insert of the vector deposited on February 5, 1998 as ATCC 209621 which includes the nucleotide sequence encoding PRO361.

In another embodiment, the invention provides isolated PRO361 polypeptide. In particular, the invention provides isolated native sequence PRO361 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 to 431 of Figure 32 (SEQ ID NO:83). An additional embodiment of the present invention is directed to an isolated extracellular domain of a PRO361 polypeptide having amino acids 1 to 379 of the amino acids sequence shown in Figure 32 (SEQ ID NO:83). Optionally, the PRO361 polypeptide is obtained or is obtainable by expressing the polypeptide encoded by the cDNA insert of the vector deposited on February 5, 1998 as ATCC 209621.

10

15. PRO365

Applicants have identified a cDNA clone that encodes a novel polypeptide, wherein the polypeptide is designated in the present application as "PRO365".

In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding a PRO365 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO365 polypeptide having amino acid residues 1 to 235 of Figure 34 (SEQ ID NO:91), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions. In another aspect, the isolated nucleic acid comprises DNA encoding the PRO365 polypeptide having amino acid residues 21 to 235 of Figure 34 (SEQ ID NO:91), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions.

In another embodiment, the invention provides isolated PRO365 polypeptide. In particular, the invention provides isolated native sequence PRO365 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 to 235 of Figure 34 (SEQ ID NO:91). An additional embodiment of the present invention is directed to an amino acid sequence comprising residues 21 to 235 of Figure 34 (SEQ ID NO:91).

16. Additional Embodiments

In other embodiments of the present invention, the invention provides vectors comprising DNA encoding any of the herein described polypeptides. Host cell comprising any such vector are also provided. By way of example, the host cells may be CHO cells, *E. coli*, or yeast. A process for producing any of the herein described polypeptides is further provided and comprises culturing host cells under conditions suitable for expression of the desired polypeptide and recovering the desired polypeptide from the cell culture.

In other embodiments, the invention provides chimeric molecules comprising any of the herein described polypeptides fused to a heterologous polypeptide or amino acid sequence. Example of such chimeric molecules comprise any of the herein described polypeptides fused to an epitope tag sequence or a Fc region of an immunoglobulin.

In another embodiment, the invention provides an antibody which specifically binds to any of the above or below described polypeptides. Optionally, the antibody is a monoclonal antibody, humanized antibody, antibody fragment or single-chain antibody.

In yet other embodiments, the invention provides oligonucleotide probes useful for isolating genomic and cDNA nucleotide sequences or as antisense probes, wherein those probes may be derived from any of the 5 above or below described nucleotide sequences.

In other embodiments, the invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a PRO polypeptide.

In one aspect, the isolated nucleic acid molecule comprises a nucleotide sequence having at least about 80% sequence identity, preferably at least about 81% sequence identity, more preferably at least about 82% 10 sequence identity, yet more preferably at least about 83% sequence identity, yet more preferably at least about 84% sequence identity, yet more preferably at least about 85% sequence identity, yet more preferably at least about 86% sequence identity, yet more preferably at least about 87% sequence identity, yet more preferably at least about 88% sequence identity, yet more preferably at least about 89% sequence identity, yet more preferably at least about 90% sequence identity, yet more preferably at least about 91% sequence identity, yet more 15 preferably at least about 92% sequence identity, yet more preferably at least about 93% sequence identity, yet more preferably at least about 94% sequence identity, yet more preferably at least about 95% sequence identity, yet more preferably at least about 96% sequence identity, yet more preferably at least about 97% sequence identity, yet more preferably at least about 98% sequence identity and yet more preferably at least about 99% sequence identity to (a) a DNA molecule encoding a PRO polypeptide having a full-length amino acid sequence as disclosed herein, an amino acid sequence lacking the signal peptide as disclosed herein, an extracellular domain of a transmembrane protein, with or without the signal peptide, as disclosed herein or any other specifically defined fragment of the full-length amino acid sequence as disclosed herein, or (b) the complement 20 of the DNA molecule of (a).

In other aspects, the isolated nucleic acid molecule comprises a nucleotide sequence having at least about 25 80% sequence identity, preferably at least about 81% sequence identity, more preferably at least about 82% sequence identity, yet more preferably at least about 83% sequence identity, yet more preferably at least about 84% sequence identity, yet more preferably at least about 85% sequence identity, yet more preferably at least about 86% sequence identity, yet more preferably at least about 87% sequence identity, yet more preferably at least about 88% sequence identity, yet more preferably at least about 89% sequence identity, yet more preferably at least about 90% sequence identity, yet more preferably at least about 91% sequence identity, yet more 30 preferably at least about 92% sequence identity, yet more preferably at least about 93% sequence identity, yet more preferably at least about 94% sequence identity, yet more preferably at least about 95% sequence identity, yet more preferably at least about 96% sequence identity, yet more preferably at least about 97% sequence identity, yet more preferably at least about 98% sequence identity and yet more preferably at least about 99% sequence identity to (a) a DNA molecule comprising the coding sequence of a full-length PRO polypeptide cDNA 35 as disclosed herein, the coding sequence of a PRO polypeptide lacking the signal peptide as disclosed herein, the coding sequence of an extracellular domain of a transmembrane PRO polypeptide, with or without the signal

peptide, as disclosed herein or the coding sequence of any other specifically defined fragment of the full-length amino acid sequence as disclosed herein, or (b) the complement of the DNA molecule of (a).

In a further aspect, the invention concerns an isolated nucleic acid molecule comprising a nucleotide sequence having at least about 80% sequence identity, preferably at least about 81% sequence identity, more preferably at least about 82% sequence identity, yet more preferably at least about 83% sequence identity, yet 5 more preferably at least about 84% sequence identity, yet more preferably at least about 85% sequence identity, yet more preferably at least about 86% sequence identity, yet more preferably at least about 87% sequence identity, yet more preferably at least about 88% sequence identity, yet more preferably at least about 89% sequence identity, yet more preferably at least about 90% sequence identity, yet more preferably at least about 91% sequence identity, yet more preferably at least about 92% sequence identity, yet more preferably at least 10 about 93% sequence identity, yet more preferably at least about 94% sequence identity, yet more preferably at least about 95% sequence identity, yet more preferably at least about 96% sequence identity, yet more preferably at least about 97% sequence identity, yet more preferably at least about 98% sequence identity and yet more preferably at least about 99% sequence identity to (a) a DNA molecule that encodes the same mature polypeptide encoded by any of the human protein cDNAs deposited with the ATCC as disclosed herein, or (b) the complement of the DNA molecule of (a).

Another aspect the invention provides an isolated nucleic acid molecule comprising a nucleotide sequence encoding a PRO polypeptide which is either transmembrane domain-deleted or transmembrane domain-inactivated, or is complementary to such encoding nucleotide sequence, wherein the transmembrane domain(s) of such polypeptide are disclosed herein. Therefore, soluble extracellular domains of the herein described PRO polypeptides are contemplated.

Another embodiment is directed to fragments of a PRO polypeptide coding sequence, or the complement thereof, that may find use as, for example, hybridization probes, for encoding fragments of a PRO polypeptide that may optionally encode a polypeptide comprising a binding site for an anti-PRO antibody or as antisense oligonucleotide probes. Such nucleic acid fragments are usually at least about 20 nucleotides in length, 25 preferably at least about 30 nucleotides in length, more preferably at least about 40 nucleotides in length, yet more preferably at least about 50 nucleotides in length, yet more preferably at least about 60 nucleotides in length, yet more preferably at least about 70 nucleotides in length, yet more preferably at least about 80 nucleotides in length, yet more preferably at least about 90 nucleotides in length, yet more preferably at least about 100 nucleotides in length, yet more preferably at least about 110 nucleotides in length, yet more preferably 30 at least about 120 nucleotides in length, yet more preferably at least about 130 nucleotides in length, yet more preferably at least about 140 nucleotides in length, yet more preferably at least about 150 nucleotides in length, yet more preferably at least about 160 nucleotides in length, yet more preferably at least about 170 nucleotides in length, yet more preferably at least about 180 nucleotides in length, yet more preferably at least about 190 nucleotides in length, yet more preferably at least about 200 nucleotides in length, yet more preferably at least 35 about 250 nucleotides in length, yet more preferably at least about 300 nucleotides in length, yet more preferably at least about 350 nucleotides in length, yet more preferably at least about 400 nucleotides in length, yet more preferably at least about 450 nucleotides in length, yet more preferably at least about 500 nucleotides in length,

yet more preferably at least about 600 nucleotides in length, yet more preferably at least about 700 nucleotides in length, yet more preferably at least about 800 nucleotides in length, yet more preferably at least about 900 nucleotides in length and yet more preferably at least about 1000 nucleotides in length, wherein in this context the term "about" means the referenced nucleotide sequence length plus or minus 10% of that referenced length. It is noted that novel fragments of a PRO polypeptide-encoding nucleotide sequence may be determined in a 5 routine manner by aligning the PRO polypeptide-encoding nucleotide sequence with other known nucleotide sequences using any of a number of well known sequence alignment programs and determining which PRO polypeptide-encoding nucleotide sequence fragment(s) are novel. All of such PRO polypeptide-encoding nucleotide sequences are contemplated herein. Also contemplated are the PRO polypeptide fragments encoded by these nucleotide molecule fragments, preferably those PRO polypeptide fragments that comprise a binding 10 site for an anti-PRO antibody.

In another embodiment, the invention provides isolated PRO polypeptide encoded by any of the isolated nucleic acid sequences hereinabove identified.

In a certain aspect, the invention concerns an isolated PRO polypeptide, comprising an amino acid sequence having at least about 80% sequence identity, preferably at least about 81% sequence identity, more preferably at least about 82% sequence identity, yet more preferably at least about 83% sequence identity, yet more preferably at least about 84% sequence identity, yet more preferably at least about 85% sequence identity, yet more preferably at least about 86% sequence identity, yet more preferably at least about 87% sequence identity, yet more preferably at least about 88% sequence identity, yet more preferably at least about 89% sequence identity, yet more preferably at least about 90% sequence identity, yet more preferably at least about 91% sequence identity, yet more preferably at least about 92% sequence identity, yet more preferably at least about 93% sequence identity, yet more preferably at least about 94% sequence identity, yet more preferably at least about 95% sequence identity, yet more preferably at least about 96% sequence identity, yet more preferably at least about 97% sequence identity, yet more preferably at least about 98% sequence identity and yet more preferably at least about 99% sequence identity to a PRO polypeptide having a full-length amino acid sequence 20 as disclosed herein, an amino acid sequence lacking the signal peptide as disclosed herein, an extracellular domain of a transmembrane protein, with or without the signal peptide, as disclosed herein or any other specifically defined fragment of the full-length amino acid sequence as disclosed herein.

In a further aspect, the invention concerns an isolated PRO polypeptide comprising an amino acid sequence having at least about 80% sequence identity, preferably at least about 81% sequence identity, more 30 preferably at least about 82% sequence identity, yet more preferably at least about 83% sequence identity, yet more preferably at least about 84% sequence identity, yet more preferably at least about 85% sequence identity, yet more preferably at least about 86% sequence identity, yet more preferably at least about 87% sequence identity, yet more preferably at least about 88% sequence identity, yet more preferably at least about 89% sequence identity, yet more preferably at least about 90% sequence identity, yet more preferably at least about 91% sequence identity, yet more preferably at least about 92% sequence identity, yet more preferably at least about 93% sequence identity, yet more preferably at least about 94% sequence identity, yet more preferably at least about 95% sequence identity, yet more preferably at least about 96% sequence identity, yet more preferably

at least about 97% sequence identity, yet more preferably at least about 98% sequence identity and yet more preferably at least about 99% sequence identity to an amino acid sequence encoded by any of the human protein cDNAs deposited with the ATCC as disclosed herein.

In a further aspect, the invention concerns an isolated PRO polypeptide comprising an amino acid sequence scoring at least about 80% positives, preferably at least about 81% positives, more preferably at least about 82% positives, yet more preferably at least about 83% positives, yet more preferably at least about 84% positives, yet more preferably at least about 85% positives, yet more preferably at least about 86% positives, yet more preferably at least about 87% positives, yet more preferably at least about 88% positives, yet more preferably at least about 89% positives, yet more preferably at least about 90% positives, yet more preferably at least about 91% positives, yet more preferably at least about 92% positives, yet more preferably at least about 93% positives, yet more preferably at least about 94% positives, yet more preferably at least about 95% positives, yet more preferably at least about 96% positives, yet more preferably at least about 97% positives, yet more preferably at least about 98% positives and yet more preferably at least about 99% positives when compared with the amino acid sequence of a PRO polypeptide having a full-length amino acid sequence as disclosed herein, an amino acid sequence lacking the signal peptide as disclosed herein, an extracellular domain of a transmembrane protein, with or without the signal peptide, as disclosed herein or any other specifically defined fragment of the full-length amino acid sequence as disclosed herein.

In a specific aspect, the invention provides an isolated PRO polypeptide without the N-terminal signal sequence and/or the initiating methionine and is encoded by a nucleotide sequence that encodes such an amino acid sequence as hereinbefore described. Processes for producing the same are also herein described, wherein those processes comprise culturing a host cell comprising a vector which comprises the appropriate encoding nucleic acid molecule under conditions suitable for expression of the PRO polypeptide and recovering the PRO polypeptide from the cell culture.

Another aspect the invention provides an isolated PRO polypeptide which is either transmembrane domain-deleted or transmembrane domain-inactivated. Processes for producing the same are also herein described, wherein those processes comprise culturing a host cell comprising a vector which comprises the appropriate encoding nucleic acid molecule under conditions suitable for expression of the PRO polypeptide and recovering the PRO polypeptide from the cell culture.

In yet another embodiment, the invention concerns agonists and antagonists of a native PRO polypeptide as defined herein. In a particular embodiment, the agonist or antagonist is an anti-PRO antibody or a small molecule.

In a further embodiment, the invention concerns a method of identifying agonists or antagonists to a PRO polypeptide which comprise contacting the PRO polypeptide with a candidate molecule and monitoring a biological activity mediated by said PRO polypeptide. Preferably, the PRO polypeptide is a native PRO polypeptide.

In a still further embodiment, the invention concerns a composition of matter comprising a PRO polypeptide, or an agonist or antagonist of a PRO polypeptide as herein described, or an anti-PRO antibody, in combination with a carrier. Optionally, the carrier is a pharmaceutically acceptable carrier.

Another embodiment of the present invention is directed to the use of a PRO polypeptide, or an agonist or antagonist thereof as hereinbefore described, or an anti-PRO antibody, for the preparation of a medicament useful in the treatment of a condition which is responsive to the PRO polypeptide, an agonist or antagonist thereof or an anti-PRO antibody.

5

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a nucleotide sequence (SEQ ID NO:1) of a native sequence PRO241 cDNA, wherein SEQ ID NO:1 is a clone designated herein as "DNA34392-1170".

Figure 2 shows the amino acid sequence (SEQ ID NO:2) derived from the coding sequence of SEQ ID NO:1 shown in Figure 1.

10 Figure 3 shows a nucleotide sequence (SEQ ID NO:6) of a native sequence PRO243 cDNA, wherein SEQ ID NO:6 is a clone designated herein as "DNA35917-1207".

Figure 4 shows the amino acid sequence (SEQ ID NO:7) derived from the coding sequence of SEQ ID NO:6 shown in Figure 3.

15 Figure 5 shows a nucleotide sequence (SEQ ID NO:14) of a native sequence PRO299 cDNA, wherein SEQ ID NO:14 is a clone designated herein as "DNA39976-1215".

Figure 6 shows the amino acid sequence (SEQ ID NO:15) derived from the coding sequence of SEQ ID NO:14 shown in Figure 5.

Figure 7 shows a nucleotide sequence designated herein as DNA28847 (SEQ ID NO:18).

Figure 8 shows a nucleotide sequence designated herein as DNA35877 (SEQ ID NO:19).

20 Figure 9 shows a nucleotide sequence (SEQ ID NO:23) of a native sequence PRO323 cDNA, wherein SEQ ID NO:23 is a clone designated herein as "DNA35595-1228".

Figure 10 shows the amino acid sequence (SEQ ID NO:24) derived from the coding sequence of SEQ ID NO:23 shown in Figure 9.

25 Figure 11 shows a single-stranded nucleotide sequence (SEQ ID NO:29) containing the nucleotide sequence (nucleotides 79-1416) of a chimeric fusion protein between a PRO323-derived polypeptide and a portion of an IgG constant domain, wherein the chimeric fusion protein is designated herein as "PRO454". The single-stranded nucleotide sequence (SEQ ID NO:29) encoding the PRO323/IgG fusion protein (PRO454) is designated herein as "DNA35872".

30 Figure 12 shows the amino acid sequence (SEQ ID NO:30) derived from nucleotides 79-1416 of the nucleotide sequence shown in Figure 11. The junction in the PRO454 amino acid sequence between the PRO323-derived sequences and the IgG-derived sequences appears between amino acids 415-416 in the figure.

Figure 13 shows a nucleotide sequence (SEQ ID NO:31) of a native sequence PRO327 cDNA, wherein SEQ ID NO:31 is a clone designated herein as "DNA38113-1230".

35 Figure 14 shows the amino acid sequence (SEQ ID NO:32) derived from the coding sequence of SEQ ID NO:31 shown in Figure 13.

Figure 15 shows a nucleotide sequence (SEQ ID NO:36) of a native sequence PRO233 cDNA, wherein SEQ ID NO:36 is a clone designated herein as "DNA34436-1238".

Figure 16 shows the amino acid sequence (SEQ ID NO:37) derived from the coding sequence of SEQ ID NO:36 shown in Figure 15.

Figure 17 shows a nucleotide sequence (SEQ ID NO:41) of a native sequence PRO344 cDNA, wherein SEQ ID NO:41 is a clone designated herein as "DNA40592-1242".

Figure 18 shows the amino acid sequence (SEQ ID NO:42) derived from the coding sequence of SEQ 5 ID NO:41 shown in Figure 17.

Figure 19 shows a nucleotide sequence (SEQ ID NO:49) of a native sequence PRO347 cDNA, wherein SEQ ID NO:49 is a clone designated herein as "DNA44176-1244".

Figure 20 shows the amino acid sequence (SEQ ID NO:50) derived from the coding sequence of SEQ ID NO:49 shown in Figure 19.

10 Figure 21 shows a nucleotide sequence (SEQ ID NO:54) of a native sequence PRO354 cDNA, wherein SEQ ID NO:54 is a clone designated herein as "DNA44192-1246".

Figure 22 shows the amino acid sequence (SEQ ID NO:55) derived from the coding sequence of SEQ ID NO:54 shown in Figure 21.

Figure 23 shows a nucleotide sequence (SEQ ID NO:60) of a native sequence PRO355 cDNA, wherein SEQ ID NO:60 is a clone designated herein as "DNA39518-1247".

Figure 24 shows the amino acid sequence (SEQ ID NO:61) derived from the coding sequence of SEQ ID NO:60 shown in Figure 23.

Figure 25 shows a nucleotide sequence (SEQ ID NO:68) of a native sequence PRO357 cDNA, wherein SEQ ID NO:68 is a clone designated herein as "DNA44804-1248".

20 Figure 26 shows the amino acid sequence (SEQ ID NO:69) derived from the coding sequence of SEQ ID NO:68 shown in Figure 25.

Figure 27 shows a nucleotide sequence (SEQ ID NO:75) of a native sequence PRO715 cDNA, wherein SEQ ID NO:75 is a clone designated herein as "DNA52722-1229".

25 Figure 28 shows the amino acid sequence (SEQ ID NO:76) derived from the coding sequence of SEQ ID NO:75 shown in Figure 27.

Figure 29 shows a nucleotide sequence (SEQ ID NO:77) of a native sequence PRO353 cDNA, wherein SEQ ID NO:77 is a clone designated herein as "DNA41234-1242".

Figure 30 shows the amino acid sequence (SEQ ID NO:78) derived from the coding sequence of SEQ ID NO:77 shown in Figure 29.

30 Figure 31 shows a nucleotide sequence (SEQ ID NO:82) of a native sequence PRO361 cDNA, wherein SEQ ID NO:82 is a clone designated herein as "DNA45410-1250".

Figure 32 shows the amino acid sequence (SEQ ID NO:83) derived from the coding sequence of SEQ ID NO:82 shown in Figure 31.

35 Figure 33 shows a nucleotide sequence (SEQ ID NO:90) of a native sequence PRO365 cDNA, wherein SEQ ID NO:90 is a clone designated herein as "DNA46777-1253".

Figure 34 shows the amino acid sequence (SEQ ID NO:91) derived from the coding sequence of SEQ ID NO:90 shown in Figure 33.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The terms "PRO polypeptide" and "PRO" as used herein and when immediately followed by a numerical designation refer to various polypeptides, wherein the complete designation (i.e., PRO/number) refers to specific polypeptide sequences as described herein. The terms "PRO/number polypeptide" and "PRO/number" wherein the term "number" is provided as an actual numerical designation as used herein 5 encompass native sequence polypeptides and polypeptide variants (which are further defined herein). The PRO polypeptides described herein may be isolated from a variety of sources, such as from human tissue types or from another source, or prepared by recombinant or synthetic methods.

A "native sequence PRO polypeptide" comprises a polypeptide having the same amino acid sequence as the corresponding PRO polypeptide derived from nature. Such native sequence PRO polypeptides can be 10 isolated from nature or can be produced by recombinant or synthetic means. The term "native sequence PRO polypeptide" specifically encompasses naturally-occurring truncated or secreted forms of the specific PRO polypeptide (e.g., an extracellular domain sequence), naturally-occurring variant forms (e.g., alternatively spliced forms) and naturally-occurring allelic variants of the polypeptide. In various embodiments of the invention, the native sequence PRO polypeptides disclosed herein are mature or full-length native sequence polypeptides comprising the full-length amino acids sequences shown in the accompanying figures. Start and stop codons are shown in bold font and underlined in the figures. However, while the PRO polypeptide disclosed in the accompanying figures are shown to begin with methionine residues designated herein as amino acid position 1 in the figures, it is conceivable and possible that other methionine residues located either upstream or downstream from the amino acid position 1 in the figures may be employed as the starting amino acid residue 15 for the PRO polypeptides.

The PRO polypeptide "extracellular domain" or "ECD" refers to a form of the PRO polypeptide which is essentially free of the transmembrane and cytoplasmic domains. Ordinarily, a PRO polypeptide ECD will have less than 1% of such transmembrane and/or cytoplasmic domains and preferably, will have less than 0.5% of such domains. It will be understood that any transmembrane domains identified for the PRO polypeptides of 20 the present invention are identified pursuant to criteria routinely employed in the art for identifying that type of hydrophobic domain. The exact boundaries of a transmembrane domain may vary but most likely by no more than about 5 amino acids at either end of the domain as initially identified herein. Optionally, therefore, an extracellular domain of a PRO polypeptide may contain from about 5 or fewer amino acids on either side of the transmembrane domain/extracellular domain boundary as identified in the Examples or specification and such 25 30 polypeptides, with or without the associated signal peptide, and nucleic acid encoding them, are contemplated by the present invention.

The approximate location of the "signal peptides" of the various PRO polypeptides disclosed herein are shown in the present specification and/or the accompanying figures. It is noted, however, that the C-terminal boundary of a signal peptide may vary, but most likely by no more than about 5 amino acids on either side of 35 the signal peptide C-terminal boundary as initially identified herein, wherein the C-terminal boundary of the signal peptide may be identified pursuant to criteria routinely employed in the art for identifying that type of amino acid sequence element (e.g., Nielsen et al., Prot. Eng. 10:1-6 (1997) and von Heinje et al., Nucl. Acids.

Res. 14:4683-4690 (1986)). Moreover, it is also recognized that, in some cases, cleavage of a signal sequence from a secreted polypeptide is not entirely uniform, resulting in more than one secreted species. These mature polypeptides, where the signal peptide is cleaved within no more than about 5 amino acids on either side of the C-terminal boundary of the signal peptide as identified herein, and the polynucleotides encoding them, are contemplated by the present invention.

5 "PRO polypeptide variant" means an active PRO polypeptide as defined above or below having at least about 80% amino acid sequence identity with a full-length native sequence PRO polypeptide sequence as disclosed herein, a PRO polypeptide sequence lacking the signal peptide as disclosed herein, an extracellular domain of a PRO polypeptide, with or without the signal peptide, as disclosed herein or any other fragment of a full-length PRO polypeptide sequence as disclosed herein. Such PRO polypeptide variants include, for 10 instance, PRO polypeptides wherein one or more amino acid residues are added, or deleted, at the N- or C-terminus of the full-length native amino acid sequence. Ordinarily, a PRO polypeptide variant will have at least about 80% amino acid sequence identity, preferably at least about 81% amino acid sequence identity, more preferably at least about 82% amino acid sequence identity, more preferably at least about 83% amino acid sequence identity, more preferably at least about 84% amino acid sequence identity, more preferably at least about 85% amino acid sequence identity, more preferably at least about 86% amino acid sequence identity, more preferably at least about 87% amino acid sequence identity, more preferably at least about 88% amino acid sequence identity, more preferably at least about 89% amino acid sequence identity, more preferably at least about 90% amino acid sequence identity, more preferably at least about 91% amino acid sequence identity, more preferably at least about 92% amino acid sequence identity, more preferably at least about 93% amino acid sequence identity, more preferably at least about 94% amino acid sequence identity, more preferably at least about 95% amino acid sequence identity, more preferably at least about 96% amino acid sequence identity, more preferably at least about 97% amino acid sequence identity, more preferably at least about 98% amino acid sequence identity and most preferably at least about 99% amino acid sequence identity with a full-length native sequence PRO polypeptide sequence as disclosed herein, a PRO polypeptide sequence lacking the signal peptide 15 as disclosed herein, an extracellular domain of a PRO polypeptide, with or without the signal peptide, as disclosed herein or any other specifically defined fragment of a full-length PRO polypeptide sequence as disclosed herein. Ordinarily, PRO variant polypeptides are at least about 10 amino acids in length, often at least about 20 amino acids in length, more often at least about 30 amino acids in length, more often at least about 40 amino acids in length, more often at least about 50 amino acids in length, more often at least about 60 amino 20 acids in length, more often at least about 70 amino acids in length, more often at least about 80 amino acids in length, more often at least about 90 amino acids in length, more often at least about 100 amino acids in length, more often at least about 150 amino acids in length, more often at least about 200 amino acids in length, more often at least about 300 amino acids in length, or more. 25

30 "Percent (%) amino acid sequence identity" with respect to the PRO polypeptide sequences identified herein is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in the specific PRO polypeptide sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative

substitutions as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST, BLAST-2, ALIGN or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared. For purposes herein, 5 however, % amino acid sequence identity values are generated using the sequence comparison computer program ALIGN-2, wherein the complete source code for the ALIGN-2 program is provided in Table 1 below. The ALIGN-2 sequence comparison computer program was authored by Genentech, Inc. and the source code shown in Table 1 below has been filed with user documentation in the U.S. Copyright Office, Washington D.C., 20559, where it is registered under U.S. Copyright Registration No. TXU510087. The ALIGN-2 program is publicly 10 available through Genentech, Inc., South San Francisco, California or may be compiled from the source code provided in Table 1 below. The ALIGN-2 program should be compiled for use on a UNIX operating system, preferably digital UNIX V4.0D. All sequence comparison parameters are set by the ALIGN-2 program and do not vary.

15 In situations where ALIGN-2 is employed for amino acid sequence comparisons, the % amino acid sequence identity of a given amino acid sequence A to, with, or against a given amino acid sequence B (which can alternatively be phrased as a given amino acid sequence A that has or comprises a certain % amino acid sequence identity to, with, or against a given amino acid sequence B) is calculated as follows:

$$100 \text{ times the fraction } X/Y$$

20 where X is the number of amino acid residues scored as identical matches by the sequence alignment program ALIGN-2 in that program's alignment of A and B, and where Y is the total number of amino acid residues in B. It will be appreciated that where the length of amino acid sequence A is not equal to the length of amino acid sequence B, the % amino acid sequence identity of A to B will not equal the % amino acid sequence identity of B to A. As examples of % amino acid sequence identity calculations using this method, Tables 2 and 3 25 demonstrate how to calculate the % amino acid sequence identity of the amino acid sequence designated "Comparison Protein" to the amino acid sequence designated "PRO", wherein "PRO" represents the amino acid sequence of a hypothetical PRO polypeptide of interest, "Comparison Protein" represents the amino acid sequence of a polypeptide against which the "PRO" polypeptide of interest is being compared, and "X", "Y" and 30 "Z" each represent different hypothetical amino acid residues.

35 Unless specifically stated otherwise, all % amino acid sequence identity values used herein are obtained as described in the immediately preceding paragraph using the ALIGN-2 computer program. However, % amino acid sequence identity values may also be obtained as described below by using the WU-BLAST-2 computer program (Altschul et al., Methods in Enzymology 266:460-480 (1996)). Most of the WU-BLAST-2 search parameters are set to the default values. Those not set to default values, i.e., the adjustable parameters, are set with the following values: overlap span = 1, overlap fraction = 0.125, word threshold (T) = 11, and scoring matrix = BLOSUM62. When WU-BLAST-2 is employed, a % amino acid sequence identity value is

determined by dividing (a) the number of matching identical amino acid residues between the amino acid sequence of the PRO polypeptide of interest having a sequence derived from the native PRO polypeptide and the comparison amino acid sequence of interest (i.e., the sequence against which the PRO polypeptide of interest is being compared which may be a PRO variant polypeptide) as determined by WU-BLAST-2 by (b) the total number of amino acid residues of the PRO polypeptide of interest. For example, in the statement "a polypeptide 5 comprising an the amino acid sequence A which has or having at least 80% amino acid sequence identity to the amino acid sequence B", the amino acid sequence A is the comparison amino acid sequence of interest and the amino acid sequence B is the amino acid sequence of the PRO polypeptide of interest.

Percent amino acid sequence identity may also be determined using the sequence comparison program NCBI-BLAST2 (Altschul et al., Nucleic Acids Res. 25:3389-3402 (1997)). The NCBI-BLAST2 sequence 10 comparison program may be downloaded from <http://www.ncbi.nlm.nih.gov>. NCBI-BLAST2 uses several search parameters, wherein all of those search parameters are set to default values including, for example, unmask = yes, strand = all, expected occurrences = 10, minimum low complexity length = 15/5, multi-pass e-value = 0.01, constant for multi-pass = 25, dropoff for final gapped alignment = 25 and scoring matrix = BLOSUM62.

In situations where NCBI-BLAST2 is employed for amino acid sequence comparisons, the % amino acid sequence identity of a given amino acid sequence A to, with, or against a given amino acid sequence B (which can alternatively be phrased as a given amino acid sequence A that has or comprises a certain % amino acid sequence identity to, with, or against a given amino acid sequence B) is calculated as follows:

20 100 times the fraction X/Y

where X is the number of amino acid residues scored as identical matches by the sequence alignment program NCBI-BLAST2 in that program's alignment of A and B, and where Y is the total number of amino acid residues in B. It will be appreciated that where the length of amino acid sequence A is not equal to the length of amino 25 acid sequence B, the % amino acid sequence identity of A to B will not equal the % amino acid sequence identity of B to A.

"PRO variant polynucleotide" or "PRO variant nucleic acid sequence" means a nucleic acid molecule which encodes an active PRO polypeptide as defined below and which has at least about 80% nucleic acid sequence identity with a nucleotide acid sequence encoding a full-length native sequence PRO polypeptide 30 sequence as disclosed herein, a full-length native sequence PRO polypeptide sequence lacking the signal peptide as disclosed herein, an extracellular domain of a PRO polypeptide, with or without the signal peptide, as disclosed herein or any other fragment of a full-length PRO polypeptide sequence as disclosed herein. Ordinarily, a PRO variant polynucleotide will have at least about 80% nucleic acid sequence identity, more preferably at least about 81% nucleic acid sequence identity, more preferably at least about 82% nucleic acid 35 sequence identity, more preferably at least about 83% nucleic acid sequence identity, more preferably at least about 84% nucleic acid sequence identity, more preferably at least about 85% nucleic acid sequence identity, more preferably at least about 86% nucleic acid sequence identity, more preferably at least about 87% nucleic

acid sequence identity, more preferably at least about 88% nucleic acid sequence identity, more preferably at least about 89% nucleic acid sequence identity, more preferably at least about 90% nucleic acid sequence identity, more preferably at least about 91% nucleic acid sequence identity, more preferably at least about 92% nucleic acid sequence identity, more preferably at least about 93% nucleic acid sequence identity, more preferably at least about 94% nucleic acid sequence identity, more preferably at least about 95% nucleic acid sequence identity, more preferably at least about 96% nucleic acid sequence identity, more preferably at least about 97% nucleic acid sequence identity, more preferably at least about 98% nucleic acid sequence identity and yet more preferably at least about 99% nucleic acid sequence identity with a nucleic acid sequence encoding a full-length native sequence PRO polypeptide sequence as disclosed herein, a full-length native sequence PRO polypeptide sequence lacking the signal peptide as disclosed herein, an extracellular domain of a PRO polypeptide, with or without the signal sequence, as disclosed herein or any other fragment of a full-length PRO polypeptide sequence as disclosed herein. Variants do not encompass the native nucleotide sequence.

Ordinarily, PRO variant polynucleotides are at least about 30 nucleotides in length, often at least about 60 nucleotides in length, more often at least about 90 nucleotides in length, more often at least about 120 nucleotides in length, more often at least about 150 nucleotides in length, more often at least about 180 nucleotides in length, more often at least about 210 nucleotides in length, more often at least about 240 nucleotides in length, more often at least about 270 nucleotides in length, more often at least about 300 nucleotides in length, more often at least about 450 nucleotides in length, more often at least about 600 nucleotides in length, more often at least about 900 nucleotides in length, or more.

"Percent (%) nucleic acid sequence identity" with respect to PRO-encoding nucleic acid sequences identified herein is defined as the percentage of nucleotides in a candidate sequence that are identical with the nucleotides in the PRO nucleic acid sequence of interest, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity. Alignment for purposes of determining percent nucleic acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST, BLAST-2, ALIGN or Megalign (DNASTAR) software. For purposes herein, however, % nucleic acid sequence identity values are generated using the sequence comparison computer program ALIGN-2, wherein the complete source code for the ALIGN-2 program is provided in Table 1 below. The ALIGN-2 sequence comparison computer program was authored by Genentech, Inc. and the source code shown in Table 1 below has been filed with user documentation in the U.S. Copyright Office, Washington D.C., 20559, where it is registered under U.S. Copyright Registration No. TXUS510087. The ALIGN-2 program is publicly available through Genentech, Inc., South San Francisco, California or may be compiled from the source code provided in Table 1 below. The ALIGN-2 program should be compiled for use on a UNIX operating system, preferably digital UNIX V4.0D. All sequence comparison parameters are set by the ALIGN-2 program and do not vary.

In situations where ALIGN-2 is employed for nucleic acid sequence comparisons, the % nucleic acid sequence identity of a given nucleic acid sequence C to, with, or against a given nucleic acid sequence D (which can alternatively be phrased as a given nucleic acid sequence C that has or comprises a certain % nucleic acid sequence identity to, with, or against a given nucleic acid sequence D) is calculated as follows:

100 times the fraction W/Z

where W is the number of nucleotides scored as identical matches by the sequence alignment program ALIGN-2 in that program's alignment of C and D, and where Z is the total number of nucleotides in D. It will be appreciated that where the length of nucleic acid sequence C is not equal to the length of nucleic acid sequence 5 D, the % nucleic acid sequence identity of C to D will not equal the % nucleic acid sequence identity of D to C. As examples of % nucleic acid sequence identity calculations, Tables 4 and 5, demonstrate how to calculate the % nucleic acid sequence identity of the nucleic acid sequence designated "Comparison DNA" to the nucleic acid sequence designated "PRO-DNA", wherein "PRO-DNA" represents a hypothetical PRO-encoding nucleic acid sequence of interest, "Comparison DNA" represents the nucleotide sequence of a nucleic acid molecule 10 against which the "PRO-DNA" nucleic acid molecule of interest is being compared, and "N", "L" and "V" each represent different hypothetical nucleotides.

Unless specifically stated otherwise, all % nucleic acid sequence identity values used herein are obtained as described in the immediately preceding paragraph using the ALIGN-2 computer program. However, % nucleic acid sequence identity values may also be obtained as described below by using the WU-BLAST-2 computer program (Altschul et al., Methods in Enzymology 266:460-480 (1996)). Most of the WU-BLAST-2 search parameters are set to the default values. Those not set to default values, i.e., the adjustable parameters, are set with the following values: overlap span = 1, overlap fraction = 0.125, word threshold (T) = 11, and scoring matrix = BLOSUM62. When WU-BLAST-2 is employed, a % nucleic acid sequence identity value is determined by dividing (a) the number of matching identical nucleotides between the nucleic acid sequence of the PRO polypeptide-encoding nucleic acid molecule of interest having a sequence derived from the native sequence PRO polypeptide-encoding nucleic acid and the comparison nucleic acid molecule of interest (i.e., the sequence against which the PRO polypeptide-encoding nucleic acid molecule of interest is being compared which may be a variant PRO polynucleotide) as determined by WU-BLAST-2 by (b) the total number of nucleotides of the PRO polypeptide-encoding nucleic acid molecule of interest. For example, in the statement "an isolated 20 nucleic acid molecule comprising a nucleic acid sequence A which has or having at least 80% nucleic acid sequence identity to the nucleic acid sequence B", the nucleic acid sequence A is the comparison nucleic acid molecule of interest and the nucleic acid sequence B is the nucleic acid sequence of the PRO polypeptide-encoding nucleic acid molecule of interest.

Percent nucleic acid sequence identity may also be determined using the sequence comparison program 30 NCBI-BLAST2 (Altschul et al., Nucleic Acids Res. 25:3389-3402 (1997)). The NCBI-BLAST2 sequence comparison program may be downloaded from <http://www.ncbi.nlm.nih.gov>. NCBI-BLAST2 uses several search parameters, wherein all of those search parameters are set to default values including, for example, unmask = yes, strand = all, expected occurrences = 10, minimum low complexity length = 15/5, multi-pass e-value = 0.01, constant for multi-pass = 25, dropoff for final gapped alignment = 25 and scoring matrix = 35 BLOSUM62.

In situations where NCBI-BLAST2 is employed for sequence comparisons, the % nucleic acid sequence identity of a given nucleic acid sequence C to, with, or against a given nucleic acid sequence D (which can

alternatively be phrased as a given nucleic acid sequence C that has or comprises a certain % nucleic acid sequence identity to, with, or against a given nucleic acid sequence D) is calculated as follows:

100 times the fraction W/Z

5 where W is the number of nucleotides scored as identical matches by the sequence alignment program NCBI-BLAST2 in that program's alignment of C and D, and where Z is the total number of nucleotides in D. It will be appreciated that where the length of nucleic acid sequence C is not equal to the length of nucleic acid sequence D, the % nucleic acid sequence identity of C to D will not equal the % nucleic acid sequence identity of D to C.

10 In other embodiments, PRO variant polynucleotides are nucleic acid molecules that encode an active PRO polypeptide and which are capable of hybridizing, preferably under stringent hybridization and wash conditions, to nucleotide sequences encoding a full-length PRO polypeptide as disclosed herein. PRO variant polypeptides may be those that are encoded by a PRO variant polynucleotide.

15 The term "positives", in the context of sequence comparison performed as described above, includes residues in the sequences compared that are not identical but have similar properties (e.g. as a result of conservative substitutions, see Table 6 below). For purposes herein, the % value of positives is determined by dividing (a) the number of amino acid residues scoring a positive value between the PRO polypeptide amino acid sequence of interest having a sequence derived from the native PRO polypeptide sequence and the comparison amino acid sequence of interest (i.e., the amino acid sequence against which the PRO polypeptide sequence is being compared) as determined in the BLOSUM62 matrix of WU-BLAST-2 by (b) the total number of amino acid residues of the PRO polypeptide of interest.

20 Unless specifically stated otherwise, the % value of positives is calculated as described in the immediately preceding paragraph. However, in the context of the amino acid sequence identity comparisons performed as described for ALIGN-2 and NCBI-BLAST-2 above, includes amino acid residues in the sequences compared that are not only identical, but also those that have similar properties. Amino acid residues that score a positive value to an amino acid residue of interest are those that are either identical to the amino acid residue of interest or are a preferred substitution (as defined in Table 6 below) of the amino acid residue of interest.

25 For amino acid sequence comparisons using ALIGN-2 or NCBI-BLAST2, the % value of positives of a given amino acid sequence A to, with, or against a given amino acid sequence B (which can alternatively be phrased as a given amino acid sequence A that has or comprises a certain % positives to, with, or against a given amino acid sequence B) is calculated as follows:

100 times the fraction X/Y

30 35 where X is the number of amino acid residues scoring a positive value as defined above by the sequence alignment program ALIGN-2 or NCBI-BLAST2 in that program's alignment of A and B, and where Y is the total number of amino acid residues in B. It will be appreciated that where the length of amino acid sequence

A is not equal to the length of amino acid sequence B, the % positives of A to B will not equal the % positives of B to A.

"Isolated," when used to describe the various polypeptides disclosed herein, means polypeptide that has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials that would typically interfere with diagnostic or therapeutic 5 uses for the polypeptide, and may include enzymes, hormones, and other proteinaceous or non-proteinaceous solutes. In preferred embodiments, the polypeptide will be purified (1) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (2) to homogeneity by SDS-PAGE under non-reducing or reducing conditions using Coomassie blue or, preferably, silver stain. Isolated polypeptide includes polypeptide *in situ* within recombinant cells, since at least one 10 component of the PRO polypeptide natural environment will not be present. Ordinarily, however, isolated polypeptide will be prepared by at least one purification step.

An "isolated" PRO polypeptide-encoding nucleic acid or other polypeptide-encoding nucleic acid is a nucleic acid molecule that is identified and separated from at least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the polypeptide-encoding nucleic acid. An isolated polypeptide-encoding nucleic acid molecule is other than in the form or setting in which it is found in nature. Isolated polypeptide-encoding nucleic acid molecules therefore are distinguished from the specific polypeptide-encoding nucleic acid molecule as it exists in natural cells. However, an isolated polypeptide-encoding nucleic acid molecule includes polypeptide-encoding nucleic acid molecules contained in cells that ordinarily express the polypeptide where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, 25 "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

The term "antibody" is used in the broadest sense and specifically covers, for example, single anti-PRO 35 monoclonal antibodies (including agonist, antagonist, and neutralizing antibodies), anti-PRO antibody compositions with polyepitopic specificity, single chain anti-PRO antibodies, and fragments of anti-PRO antibodies (see below). The term "monoclonal antibody" as used herein refers to an antibody obtained from a

population of substantially homogeneous antibodies, i.e., the individual antibodies comprising the population are identical except for possible naturally-occurring mutations that may be present in minor amounts.

"Stringency" of hybridization reactions is readily determinable by one of ordinary skill in the art, and generally is an empirical calculation dependent upon probe length, washing temperature, and salt concentration. In general, longer probes require higher temperatures for proper annealing, while shorter probes need lower 5 temperatures. Hybridization generally depends on the ability of denatured DNA to reanneal when complementary strands are present in an environment below their melting temperature. The higher the degree of desired homology between the probe and hybridizable sequence, the higher the relative temperature which can be used. As a result, it follows that higher relative temperatures would tend to make the reaction conditions more stringent, while lower temperatures less so. For additional details and explanation of stringency of 10 hybridization reactions, see Ausubel et al., Current Protocols in Molecular Biology, Wiley Interscience Publishers, (1995).

"Stringent conditions" or "high stringency conditions", as defined herein, may be identified by those that: (1) employ low ionic strength and high temperature for washing, for example 0.015 M sodium chloride/0.0015 M sodium citrate/0.1% sodium dodecyl sulfate at 50°C; (2) employ during hybridization a denaturing agent, such as formamide, for example, 50% (v/v) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride, 75 mM sodium citrate at 42°C; or (3) employ 50% formamide, 5 x SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 µg/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 x SSC (sodium chloride/sodium citrate) and 50% formamide at 55°C, followed by a high-stringency wash 15 consisting of 0.1 x SSC containing EDTA at 55°C.

"Moderately stringent conditions" may be identified as described by Sambrook et al., Molecular Cloning: A Laboratory Manual, New York: Cold Spring Harbor Press, 1989, and include the use of washing 20 solution and hybridization conditions (e.g., temperature, ionic strength and %SDS) less stringent than those described above. An example of moderately stringent conditions is overnight incubation at 37°C in a solution comprising: 20% formamide, 5 x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5 x Denhardt's solution, 10% dextran sulfate, and 20 mg/ml denatured sheared salmon sperm DNA, followed by washing the filters in 1 x SSC at about 37-50°C. The skilled artisan will recognize how to adjust 25 the temperature, ionic strength, etc. as necessary to accommodate factors such as probe length and the like.

30 The term "epitope tagged" when used herein refers to a chimeric polypeptide comprising a PRO polypeptide fused to a "tag polypeptide". The tag polypeptide has enough residues to provide an epitope against which an antibody can be made, yet is short enough such that it does not interfere with activity of the polypeptide to which it is fused. The tag polypeptide preferably also is fairly unique so that the antibody does not substantially cross-react with other epitopes. Suitable tag polypeptides generally have at least six amino acid 35 residues and usually between about 8 and 50 amino acid residues (preferably, between about 10 and 20 amino acid residues).

As used herein, the term "immunoadhesin" designates antibody-like molecules which combine the binding specificity of a heterologous protein (an "adhesin") with the effector functions of immunoglobulin constant domains. Structurally, the immunoadhesins comprise a fusion of an amino acid sequence with the desired binding specificity which is other than the antigen recognition and binding site of an antibody (i.e., is "heterologous"), and an immunoglobulin constant domain sequence. The adhesin part of an immunoadhesin 5 molecule typically is a contiguous amino acid sequence comprising at least the binding site of a receptor or a ligand. The immunoglobulin constant domain sequence in the immunoadhesin may be obtained from any immunoglobulin, such as IgG-1, IgG-2, IgG-3, or IgG-4 subtypes, IgA (including IgA-1 and IgA-2), IgE, IgD or IgM.

"Active" or "activity" for the purposes herein refers to form(s) of a PRO polypeptide which retain a 10 biological and/or an immunological activity of native or naturally-occurring PRO, wherein "biological" activity refers to a biological function (either inhibitory or stimulatory) caused by a native or naturally-occurring PRO other than the ability to induce the production of an antibody against an antigenic epitope possessed by a native or naturally-occurring PRO and an "immunological" activity refers to the ability to induce the production of an antibody against an antigenic epitope possessed by a native or naturally-occurring PRO.

The term "antagonist" is used in the broadest sense, and includes any molecule that partially or fully blocks, inhibits, or neutralizes a biological activity of a native PRO polypeptide disclosed herein. In a similar manner, the term "agonist" is used in the broadest sense and includes any molecule that mimics a biological activity of a native PRO polypeptide disclosed herein. Suitable agonist or antagonist molecules specifically include agonist or antagonist antibodies or antibody fragments, fragments or amino acid sequence variants of native PRO polypeptides, peptides, antisense oligonucleotides, small organic molecules, etc. Methods for identifying agonists or antagonists of a PRO polypeptide may comprise contacting a PRO polypeptide with a candidate agonist or antagonist molecule and measuring a detectable change in one or more biological activities normally associated with the PRO polypeptide.

"Treatment" refers to both therapeutic treatment and prophylactic or preventative measures, wherein 25 the object is to prevent or slow down (lessen) the targeted pathologic condition or disorder. Those in need of treatment include those already with the disorder as well as those prone to have the disorder or those in whom the disorder is to be prevented.

"Chronic" administration refers to administration of the agent(s) in a continuous mode as opposed to an acute mode, so as to maintain the initial therapeutic effect (activity) for an extended period of time. 30 "Intermittent" administration is treatment that is not consecutively done without interruption, but rather is cyclic in nature.

"Mammal" for purposes of treatment refers to any animal classified as a mammal, including humans, domestic and farm animals, and zoo, sports, or pet animals, such as dogs, cats, cattle, horses, sheep, pigs, goats, rabbits, etc. Preferably, the mammal is human.

35 Administration "in combination with" one or more further therapeutic agents includes simultaneous (concurrent) and consecutive administration in any order.

"Carriers" as used herein include pharmaceutically acceptable carriers, excipients, or stabilizers which are nontoxic to the cell or mammal being exposed thereto at the dosages and concentrations employed. Often the physiologically acceptable carrier is an aqueous pH buffered solution. Examples of physiologically acceptable carriers include buffers such as phosphate, citrate, and other organic acids; antioxidants including ascorbic acid; low molecular weight (less than about 10 residues) polypeptide; proteins, such as serum albumin, 5 gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, arginine or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; salt-forming counterions such as sodium; and/or nonionic surfactants such as TWEEN™, polyethylene glycol (PEG), and PLURONICS™.

10 "Antibody fragments" comprise a portion of an intact antibody, preferably the antigen binding or variable region of the intact antibody. Examples of antibody fragments include Fab, Fab', F(ab')₂, and Fv fragments; diabodies; linear antibodies (Zapata et al., Protein Eng. 8(10): 1057-1062 [1995]); single-chain antibody molecules; and multispecific antibodies formed from antibody fragments.

15 Papain digestion of antibodies produces two identical antigen-binding fragments, called "Fab" fragments, each with a single antigen-binding site, and a residual "Fc" fragment, a designation reflecting the ability to crystallize readily. Pepsin treatment yields an F(ab')₂ fragment that has two antigen-combining sites and is still capable of cross-linking antigen.

20 "Fv" is the minimum antibody fragment which contains a complete antigen-recognition and -binding site. This region consists of a dimer of one heavy- and one light-chain variable domain in tight, non-covalent association. It is in this configuration that the three CDRs of each variable domain interact to define an antigen-binding site on the surface of the V_H-V_L dimer. Collectively, the six CDRs confer antigen-binding specificity to the antibody. However, even a single variable domain (or half of an Fv comprising only three CDRs specific for an antigen) has the ability to recognize and bind antigen, although at a lower affinity than the entire binding site.

25 The Fab fragment also contains the constant domain of the light chain and the first constant domain (CH1) of the heavy chain. Fab fragments differ from Fab' fragments by the addition of a few residues at the carboxy terminus of the heavy chain CH1 domain including one or more cysteines from the antibody hinge region. Fab'-SH is the designation herein for Fab' in which the cysteine residue(s) of the constant domains bear a free thiol group. F(ab')₂ antibody fragments originally were produced as pairs of Fab' fragments which have 30 hinge cysteines between them. Other chemical couplings of antibody fragments are also known.

The "light chains" of antibodies (immunoglobulins) from any vertebrate species can be assigned to one of two clearly distinct types, called kappa and lambda, based on the amino acid sequences of their constant domains.

35 Depending on the amino acid sequence of the constant domain of their heavy chains, immunoglobulins can be assigned to different classes. There are five major classes of immunoglobulins: IgA, IgD, IgE, IgG, and IgM, and several of these may be further divided into subclasses (isotypes), e.g., IgG1, IgG2, IgG3, IgG4, IgA, and IgA2.

"Single-chain Fv" or "sFv" antibody fragments comprise the V_H and V_L domains of antibody, wherein these domains are present in a single polypeptide chain. Preferably, the Fv polypeptide further comprises a polypeptide linker between the V_H and V_L domains which enables the sFv to form the desired structure for antigen binding. For a review of sFv, see Pluckthun in The Pharmacology of Monoclonal Antibodies, vol. 113, Rosenburg and Moore eds., Springer-Verlag, New York, pp. 269-315 (1994).

5 The term "diabodies" refers to small antibody fragments with two antigen-binding sites, which fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) in the same polypeptide chain (V_H-V_L). By using a linker that is too short to allow pairing between the two domains on the same chain, the domains are forced to pair with the complementary domains of another chain and create two antigen-binding sites. Diabodies are described more fully in, for example, EP 404,097; WO 93/11161; and
10 Hollinger et al., Proc. Natl. Acad. Sci. USA, 90:6444-6448 (1993).

An "isolated" antibody is one which has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials which would interfere with diagnostic or therapeutic uses for the antibody, and may include enzymes, hormones, and other proteinaceous or nonproteinaceous solutes. In preferred embodiments, the antibody will be purified (1) to greater than 95% by weight of antibody as determined by the Lowry method, and most preferably more than 99% by weight, (2) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (3) to homogeneity by SDS-PAGE under reducing or nonreducing conditions using Coomassie blue or, preferably, silver stain. Isolated antibody includes the antibody in situ within recombinant cells since at least one component of the antibody's natural environment will not be present. Ordinarily, however, isolated antibody will be prepared by at least one purification step.

20 The word "label" when used herein refers to a detectable compound or composition which is conjugated directly or indirectly to the antibody so as to generate a "labeled" antibody. The label may be detectable by itself (e.g. radioisotope labels or fluorescent labels) or, in the case of an enzymatic label, may catalyze chemical alteration of a substrate compound or composition which is detectable.

25 By "solid phase" is meant a non-aqueous matrix to which the antibody of the present invention can adhere. Examples of solid phases encompassed herein include those formed partially or entirely of glass (e.g., controlled pore glass), polysaccharides (e.g., agarose), polyacrylamides, polystyrene, polyvinyl alcohol and silicones. In certain embodiments, depending on the context, the solid phase can comprise the well of an assay plate; in others it is a purification column (e.g., an affinity chromatography column). This term also includes
30 a discontinuous solid phase of discrete particles, such as those described in U.S. Patent No. 4,275,149.

A "liposome" is a small vesicle composed of various types of lipids, phospholipids and/or surfactant which is useful for delivery of a drug (such as a PRO polypeptide or antibody thereto) to a mammal. The components of the liposome are commonly arranged in a bilayer formation, similar to the lipid arrangement of biological membranes.

35 A "small molecule" is defined herein to have a molecular weight below about 500 Daltons.

Table 1

40

45

50

55

Table 1 (cont')

```

/*
 */
#include <stdio.h>
#include <ctype.h>
5
#define MAXJMP 16 /* max jumps in a diag */
#define MAXGAP 24 /* don't continue to penalize gaps larger than this */
#define J MPS 1024 /* max jmps in an path */
#define MX 4 /* save if there's at least MX-1 bases since last jmp */
10
#define DMAT 3 /* value of matching bases */
#define DMIS 0 /* penalty for mismatched bases */
#define DINS0 8 /* penalty for a gap */
#define DINS1 1 /* penalty per base */
15
#define PINS0 8 /* penalty for a gap */
#define PINS1 4 /* penalty per residue */

20
struct jmp {
    short n[MAXJMP]; /* size of jmp (neg for delay) */
    unsigned short x[MAXJMP]; /* base no. of jmp in seq x */
    /* limits seq to 2^16 - 1 */
};

25
struct diag {
    int score; /* score at last jmp */
    long offset; /* offset of prev block */
    short ijmp; /* current jmp index */
    struct jmp jp; /* list of jmps */
};

30
struct path {
    int spc; /* number of leading spaces */
    short n[J MPS]; /* size of jmp (gap) */
    int x[J MPS]; /* loc of jmp (last elem before gap) */
};

35
char *ofile; /* output file name */
char *namex[2]; /* seq names: getseqs() */
char *prog; /* prog name for err msgs */
char *seqx[2]; /* seqs: getseqs() */

40
int dmax; /* best diag: nw() */
int dmax0; /* final diag */
int dna; /* set if dna: main() */
int endgaps; /* set if penalizing end gaps */
int gapx, gapy; /* total gaps in seqs */
45
int len0, len1; /* seq lens */
int ngapx, ngapy; /* total size of gaps */
int smax; /* max score: nw() */
int *xbm; /* bitmap for matching */
long offset; /* current offset in jmp file */
50
struct diag *dx; /* holds diagonals */
struct path pp[2]; /* holds path for seqs */

55
char *calloc(), *malloc(), *index(), *strcpy();
char *getseq(), *g_calloc();

```

Table 1 (cont')

```

/* Needleman-Wunsch alignment program
*
* usage: progs file1 file2
* where file1 and file2 are two dna or two protein sequences.
5 * The sequences can be in upper- or lower-case an may contain ambiguity
* Any lines beginning with ';' or '>' are ignored
* Max file length is 65535 (limited by unsigned short x in the jmp struct)
* A sequence with 1/3 or more of its elements ACGTU is assumed to be DNA
* Output is in the file "align.out"
10 *
* The program may create a tmp file in /tmp to hold info about traceback.
* Original version developed under BSD 4.3 on a vax 8650
*/
#include "nw.h"
15 #include "day.h"

static _dbval[26] = {
    1,14,2,13,0,0,4,11,0,0,12,0,3,15,0,0,0,5,6,8,8,7,9,0,10,0
};

20 static _pbval[26] = {
    1, 2|(1<<('D'-'A'))|(1<<('N'-'A')), 4, 8, 16, 32, 64,
    128, 256, 0xFFFFFFFF, 1<<10, 1<<11, 1<<12, 1<<13, 1<<14,
    1<<15, 1<<16, 1<<17, 1<<18, 1<<19, 1<<20, 1<<21, 1<<22,
    1<<23, 1<<24, 1<<25|(1<<('E'-'A'))|(1<<('Q'-'A'))
};

main(ac, av)
30     int ac;
     char *av[];
{
    prog = av[0];
    if (ac != 3) {
        fprintf(stderr, "usage: %s file1 file2\n", prog);
        fprintf(stderr, "where file1 and file2 are two dna or two protein sequences.\n");
        fprintf(stderr, "The sequences can be in upper- or lower-case\n");
        fprintf(stderr, "Any lines beginning with ';' or '>' are ignored\n");
        fprintf(stderr, "Output is in the file \"align.out\"\n");
        exit(1);
    }
40    namex[0] = av[1];
    namex[1] = av[2];
    seqx[0] = getseq(namex[0], &len0);
    seqx[1] = getseq(namex[1], &len1);
    xbm = (dna)? _dbval : _pbval;

    endgaps = 0; /* 1 to penalize endgaps */
    ofile = "align.out"; /* output file */

50    nw(); /* fill in the matrix, get the possible jmps */
    readjmps(); /* get the actual jmps */
    print(); /* print stats, alignment */

    cleanup(0); /* unlink any tmp files */
55 }

```

Table 1 (cont')

```

/* do the alignment, return best score: main()
 * dna: values in Fitch and Smith, PNAS, 80, 1382-1386, 1983
 * pro: PAM 250 values
 * When scores are equal, we prefer mismatches to any gap, prefer
5   * a new gap to extending an ongoing gap, and prefer a gap in seqx
 * to a gap in seq y.
 */
nw()
{
10    char      *px, *py;          /* seqs and ptrs */
    int       *ndely, *dely;    /* keep track of dely */
    int       ndlx, delx;      /* keep track of delx */
    int       *tmp;             /* for swapping row0, row1 */
    int       mis;              /* score for each type */
15    int       ins0, ins1;      /* insertion penalties */
    register  id;              /* diagonal index */
    register  ij;              /* jmp index */
    register  *col0, *col1;    /* score for curr, last row */
    register  xx, yy;          /* index into seqs */
20
    dx = (struct diag *)g_calloc("to get diags", len0+len1+1, sizeof(struct diag));
25
    ndely = (int *)g_calloc("to get ndely", len1+1, sizeof(int));
    dely = (int *)g_calloc("to get dely", len1+1, sizeof(int));
    col0 = (int *)g_calloc("to get col0", len1+1, sizeof(int));
    col1 = (int *)g_calloc("to get col1", len1+1, sizeof(int));
    ins0 = (dna)? DINS0 : PINS0;
    ins1 = (dna)? DINS1 : PINS1;
30
    smax = -10000;
    if (endgaps) {
        for (col0[0] = dely[0] = -ins0, yy = 1; yy <= len1; yy++) {
            col0[yy] = dely[yy] = col0[yy-1] - ins1;
            ndely[yy] = yy;
        }
        col0[0] = 0;      /* Waterman Bull Math Biol 84 */
    }
35
    else
        for (yy = 1; yy <= len1; yy++)
            dely[yy] = -ins0;
40
    /* fill in match matrix
     */
45    for (px = seqx[0], xx = 1; xx <= len0; px++, xx++) {
        /* initialize first entry in col
         */
        if (endgaps) {
            if (xx == 1)
                col1[0] = delx = -(ins0+ins1);
            else
                col1[0] = delx = col0[0] - ins1;
                ndlx = xx;
        }
50
        else {
            col1[0] = 0;
            delx = -ins0;
            ndlx = 0;
        }
55
    }
}

```

Table 1 (cont')

...nw

```
for (py = seqx[1], yy = 1; yy <= len1; py++, yy++) {
    mis = col0[yy-1];
    if (dna)
        mis += (xbm[*px-'A']&xbm[*py-'A'])? DMAT : DMIS;
    else
        mis += _day[*px-'A'][*py-'A'];

    /* update penalty for del in x seq;
     * favor new del over ongong del
     * ignore MAXGAP if weighting endgaps
     */
    if (endgaps || ndely[yy] < MAXGAP) {
        if (col0[yy] - ins0 >= dely[yy]) {
            dely[yy] = col0[yy] - (ins0+ins1);
            ndely[yy] = 1;
        } else {
            dely[yy] -= ins1;
            ndely[yy]++;
        }
    } else {
        if (col0[yy] - (ins0+ins1) >= dely[yy]) {
            dely[yy] = col0[yy] - (ins0+ins1);
            ndely[yy] = 1;
        } else
            ndely[yy]++;
    }

    /* update penalty for del in y seq;
     * favor new del over ongong del
     */
    if (endgaps || ndelx < MAXGAP) {
        if (col1[yy-1] - ins0 >= delx) {
            delx = col1[yy-1] - (ins0+ins1);
            ndelx = 1;
        } else {
            delx -= ins1;
            ndelx++;
        }
    } else {
        if (col1[yy-1] - (ins0+ins1) >= delx) {
            delx = col1[yy-1] - (ins0+ins1);
            ndelx = 1;
        } else
            ndelx++;
    }

    /* pick the maximum score; we're favoring
     * mis over any del and delx over dely
     */
}
```

55

60

Table 1 (cont')

...nw

```
id = xx - yy + len1 - 1;
if (mis >= delx && mis >= dely[yy])
    col1[yy] = mis;
5    else if (delx >= dely[yy]) {
        col1[yy] = delx;
        ij = dx[id].ijmp;
        if (dx[id].jp.n[0] && (!dna || (ndelx >= MAXJMP
&& xx > dx[id].jp.x[ij]+MX) || mis > dx[id].score+DINS0)) {
            dx[id].ijmp++;
            if (++ij >= MAXJMP) {
                writejmps(id);
                ij = dx[id].ijmp = 0;
                dx[id].offset = offset;
                offset += sizeof(struct jmp) + sizeof(offset);
            }
        }
        dx[id].jp.n[ij] = ndelx;
        dx[id].jp.x[ij] = xx;
        dx[id].score = delx;
20    }
    else {
        col1[yy] = dely[yy];
        ij = dx[id].ijmp;
        if (dx[id].jp.n[0] && (!dna || (ndely[yy] >= MAXJMP
&& xx > dx[id].jp.x[ij]+MX) || mis > dx[id].score+DINS0)) {
            dx[id].ijmp++;
            if (++ij >= MAXJMP) {
                writejmps(id);
                ij = dx[id].ijmp = 0;
                dx[id].offset = offset;
                offset += sizeof(struct jmp) + sizeof(offset);
            }
        }
        dx[id].jp.n[ij] = -ndely[yy];
        dx[id].jp.x[ij] = xx;
        dx[id].score = dely[yy];
    }
    if (xx == len0 && yy < len1) {
40        /* last col
        */
        if (endgaps)
            col1[yy] -= ins0+ins1*(len1-yy);
        if (col1[yy] > smax) {
            smax = col1[yy];
            dmax = id;
        }
    }
50    if (endgaps && xx < len0)
        col1[yy-1] -= ins0+ins1*(len0-xx);
    if (col1[yy-1] > smax) {
        smax = col1[yy-1];
        dmax = id;
    }
55    tmp = col0; col0 = col1; col1 = tmp;
}
(void) free((char *)ndely);
(void) free((char *)dely);
60 (void) free((char *)col0);
(void) free((char *)col1);
}
```

Table 1 (cont')

```
/*
*
* print() -- only routine visible outside this module
*
5  * static:
* getmat() -- trace back best path, count matches: print()
* pr_align() -- print alignment of described in array p[]: print()
* dumpblock() -- dump a block of lines with numbers, stars: pr_align()
* nums() -- put out a number line: dumpblock()
* 10  * putline() -- put out a line (name, [num], seq, [num]): dumpblock()
* stars() - -put a line of stars: dumpblock()
* stripname() -- strip any path and prefix from a seqname
*/
15 #include "nw.h"

#define SPC      3
#define P_LINE   256      /* maximum output line */
#define P_SPC    3      /* space between name or num and seq */
20
extern _day[26][26];
int olen;           /* set output line length */
FILE *fx;           /* output file */

30
print()
{
    int lx, ly, firstgap, lastgap; /* overlap */

    if ((fx = fopen(ofile, "w")) == 0) {
        fprintf(stderr, "%s: can't write %s\n", prog, ofile);
        cleanup(1);
    }
    35  fprintf(fx, "<first sequence: %s (length = %d)\n", namex[0], len0);
    fprintf(fx, "<second sequence: %s (length = %d)\n", namex[1], len1);
    olen = 60;
    lx = len0;
    ly = len1;
    firstgap = lastgap = 0;
    40  if (dmax < len1 - 1) { /* leading gap in x */
        pp[0].spc = firstgap = len1 - dmax - 1;
        ly -= pp[0].spc;
    }
    else if (dmax > len1 - 1) { /* leading gap in y */
        pp[1].spc = firstgap = dmax - (len1 - 1);
        lx -= pp[1].spc;
    }
    45  if (dmax0 < len0 - 1) { /* trailing gap in x */
        lastgap = len0 - dmax0 - 1;
        lx -= lastgap;
    }
    else if (dmax0 > len0 - 1) { /* trailing gap in y */
        lastgap = dmax0 - (len0 - 1);
        ly -= lastgap;
    }
    50  }
    55  getmat(lx, ly, firstgap, lastgap);
    pr_align();
}

60
```

Table 1 (cont')

```
/*
 * trace back the best path, count matches
 */
static
5  getmat(lx, ly, firstgap, lastgap)                                getmat
    int      lx, ly;          /* "core" (minus endgaps) */
    int      firstgap, lastgap; /* leading/trailing overlap */
{
10    int      nm, i0, i1, siz0, siz1;
    char     outx[32];
    double   pct;
    register int n0, n1;
    register char *p0, *p1;

15    /* get total matches, score
 */
    i0 = i1 = siz0 = siz1 = 0;
    p0 = seqx[0] + pp[1].spc;
    p1 = seqx[1] + pp[0].spc;
20    n0 = pp[1].spc + 1;
    n1 = pp[0].spc + 1;

25    nm = 0;
    while (*p0 && *p1) {
        if (siz0) {
30        p1++;
        n1++;
        siz0--;
        }
        else if (siz1) {
35        p0++;
        n0++;
        siz1--;
        }
        else {
40        if (xbm[*p0-'A']&xbm[*p1-'A'])
            nm++;
        if (n0++ == pp[0].x[i0])
            siz0 = pp[0].n[i0++];
        if (n1++ == pp[1].x[i1])
            siz1 = pp[1].n[i1++];
        p0++;
        p1++;
        }
45    }
    /* pct homology:
     * if penalizing endgaps, base is the shorter seq
     * else, knock off overhangs and take shorter core
50    */
    if (endgaps)
        lx = (len0 < len1)? len0 : len1;
    else
55        lx = (lx < ly)? lx : ly;
    pct = 100.*(double)nm/(double)lx;
    fprintf(fx, "\n");
    fprintf(fx, "<%d match%s in an overlap of %d: %.2f percent similarity\n",
            nm, (nm == 1)? ":" : "es", lx, pct);
```

Table 1 (cont')

```
5   fprintf(fx, "<gaps in first sequence: %d", gapx);           ...getmat
  if (gapx) {
    (void) sprintf(outx, " (%d %s%s)",                               10
      ngapx, (dna)? "base":"residue", (ngapx == 1)? ":"s");
    fprintf(fx, "%s", outx);

10  fprintf(fx, ", gaps in second sequence: %d", gapy);
  if (gapy) {
    (void) sprintf(outx, " (%d %s%s)",                               15
      ngapy, (dna)? "base":"residue", (ngapy == 1)? ":"s");
    fprintf(fx, "%s", outx);
  }
  if (dna)
    fprintf(fx, "\n<score: %d (match = %d, mismatch = %d, gap penalty = %d + %d per base)\n",
      smax, DMAT, DMIS, DINSO, DINS1);
  else
    fprintf(fx, "\n<score: %d (Dayhoff PAM 250 matrix, gap penalty = %d + %d per residue)\n",
      smax, PINS0, PINS1);
20  if (endgaps)
    fprintf(fx,
      "<endgaps penalized. left endgap: %d %s%s, right endgap: %d %s%s\n",
      firstgap, (dna)? "base" : "residue", (firstgap == 1)? ":"s",
      lastgap, (dna)? "base" : "residue", (lastgap == 1)? ":"s");
  else
    fprintf(fx, "<endgaps not penalized\n");
30  }

35  static nm;           /* matches in core -- for checking */
  static lmax;          /* lengths of stripped file names */
  static ij[2];          /* jmp index for a path */
  static nc[2];          /* number at start of current line */
  static ni[2];          /* current elem number -- for gapping */
  static siz[2];          /* */
  static *ps[2];          /* ptr to current element */
  static *po[2];          /* ptr to next output char slot */
  static out[2][P_LINE];  /* output line */
  static star[P_LINE];   /* set by stars0 */

40  /*

45  * print alignment of described in struct path pp[]
  */
45  static pr_align()           pr_align
{                                50
  int nn;           /* char count */
  int more;
  register i;

50  for (i = 0, lmax = 0; i < 2; i++) {
    nn = stripname(namex[i]);
    if (nn > lmax)
      lmax = nn;

55  nc[i] = 1;
    ni[i] = 1;
    siz[i] = ij[i] = 0;
    ps[i] = seqx[i];
    po[i] = out[i];
60  }

  }
```

Table 1 (cont')

```
for (nn = nm = 0, more = 1; more; ) {
    for (i = more = 0; i < 2; i++) {
        /*
         * do we have more of this sequence?
         */
        if (!*ps[i])
            continue;
5        more++;

10       if (pp[i].spc) { /* leading space */
            *po[i]++ = ' ';
            pp[i].spc--;
        }
15       else if (siz[i]) { /* in a gap */
            *po[i]++ = '-';
            siz[i]--;
        }
20       else { /* we're putting a seq element
            /*
             * are we at next gap for this seq?
             */
            if (ni[i] == pp[i].x[ij[i]]) {
                /*
                 * we need to merge all gaps
                 * at this location
                 */
                siz[i] = pp[i].n[ij[i]++];
                while (ni[i] == pp[i].x[ij[i]])
                    siz[i] += pp[i].n[ij[i]++];
            }
30            ni[i]++;
40        }
        }
    }
35       if (++nn == olen || !more && nn) {
45           dumpblock();
            for (i = 0; i < 2; i++)
                po[i] = out[i];
            nn = 0;
        }
50    }
}
55    /*
     * dump a block of lines, including numbers, stars: pr_align()
     */
static
dumpblock()
{
    register i;
60    for (i = 0; i < 2; i++)
        *po[i]-- = '\0';
}
...pr_align
dumpblock
```

Table 1 (cont')

```
...dumpblock

5   (void) putc('\n', fx);
  for (i = 0; i < 2; i++) {
    if (*out[i] && (*out[i] != ' ' || *(po[i]) != ' ')) {
      if (i == 0)
        nums(i);
      if (i == 0 && *out[1])
        stars();
10    putline(i);
      if (i == 0 && *out[1])
        fprintf(fx, star);
      if (i == 1)
        nums(i);
15    }
  }
}

20  /*
 * put out a number line: dumpblock()
 */
static
25  nums(ix)
  int      ix;      /* index in out[] holding seq line */
{
  char      nline[P_LINE];
  register  i, j;
  register char  *pn, *px, *py;

30  for (pn = nline, i = 0; i < lmax+P_SPC; i++, pn++)
    *pn = ' ';
  for (i = nc[ix], py = out[ix]; *py; py++, pn++) {
    if (*py == ' ' || *py == '-')
      *pn = ' ';
    else {
      if (i%10 == 0 || (i == 1 && nc[ix] != 1)) {
        j = (i < 0)? -i : i;
        for (px = pn; j /= 10, px--)
          *px = j%10 + '0';
40      if (i < 0)
          *px = '-';
      }
      else
        *pn = ' ';
    }
    i++;
45  }
  *pn = '\0';
  nc[ix] = i;
50  for (pn = nline; *pn; pn++)
    (void) putc(*pn, fx);
  (void) putc('\n', fx);
}

55  /*
 * put out a line (name, [num], seq, [num]): dumpblock()
 */
static
60  putline(ix)
  int      ix;
{
```

nums

putline

Table 1 (cont')

```
5           int      i;
register char *px;
for (px = namex[ix], i = 0; *px && *px != ':'; px++, i++)
    (void) putc(*px, fx);
for (; i < lmax + P_SPC; i++)
    (void) putc(' ', fx);
10
/* these count from 1:
 * ni[] is current element (from 1)
 * nc[] is number at start of current line
 */
15
for (px = out[ix]; *px; px++)
    (void) putc(*px&0x7F, fx);
(void) putc('\n', fx);
}
20
/*
 * put a line of stars (seqs always in out[0], out[1]): dumpblock()
 */
static
stars()
{
    int      i;
register char *p0, *p1, cx, *px;
30
if (!*out[0] || (*out[0] == ' ' && *(po[0]) == ' ') ||
    !*out[1] || (*out[1] == ' ' && *(po[1]) == ' '))
    return;
px = star;
for (i = lmax + P_SPC; i; i--)
    *px++ = ' ';
35
for (p0 = out[0], p1 = out[1]; *p0 && *p1; p0++, p1++) {
    if (isalpha(*p0) && isalpha(*p1)) {
40
        if (xbm[*p0-'A']&xbm[*p1-'A']) {
            cx = '*';
            nm++;
        }
        else if (!dna && _day[*p0-'A'][*p1-'A'] > 0)
            cx = ':';
        else
            cx = ' ';
    }
    else
50
        cx = ' ';
    *px++ = cx;
}
*px++ = '\n';
*px = '\0';
55
}
60
```

Table 1 (cont')

```
/*
 * strip path or prefix from pn, return len: pr_align()
 */
static
5  stripname(pn)
    char    *pn;      /* file name (may be path) */
{
    register char    *px, *py;

10    py = 0;
    for (px = pn; *px; px++)
        if (*px == '/')
            py = px + 1;
    if (py)
15        (void) strcpy(pn, py);
    return(strlen(pn));
}
20
```

25
30
35
40

45

50

55

60

stripname

Table 1 (cont')

```
/*
 * cleanup() -- cleanup any tmp file
 * getseq() -- read in seq, set dna, len, maxlen
 * g_calloc() -- calloc() with error checkin
5  * readjmps() -- get the good jmps, from tmp file if necessary
 * writejmps() -- write a filled array of jmps to a tmp file: nw()
 */
#include "nw.h"
#include <sys/file.h>
10
char *jname = "/tmp/homgXXXXXX";           /* tmp file for jmps */
FILE *fj;
15
int cleanup();                         /* cleanup tmp file */
long lseek();
20
/*
 * remove any tmp file if we blow
 */
cleanup(i)
25
    int i;
{
    if (fj)
        (void) unlink(jname);
    exit(i);
}
30
/*
 * read, return ptr to seq, set dna, len, maxlen
 * skip lines starting with ';', '<', or '>'
 * seq in upper or lower case
 */
35
char *
getseq(file, len)
40
    char *file; /* file name */
    int *len; /* seq len */
{
    char line[1024], *pseq;
    register char *px, *py;
    int natgc, tlen;
    FILE *fp;
45
    if ((fp = fopen(file, "r")) == 0) {
        fprintf(stderr, "%s: can't read %s\n", prog, file);
        exit(1);
    }
    tlen = natgc = 0;
    while (fgets(line, 1024, fp)) {
50
        if (*line == ';' || *line == '<' || *line == '>')
            continue;
        for (px = line; *px != '\n'; px++)
            if (isupper(*px) || islower(*px))
                tlen++;
    }
55
    if ((pseq = malloc((unsigned)(tlen+6))) == 0) {
        fprintf(stderr, "%s: malloc() failed to get %d bytes for %s\n", prog, tlen+6, file);
        exit(1);
    }
    pseq[0] = pseq[1] = pseq[2] = pseq[3] = '\0';
60
```

Table 1 (cont')

```

...getseq
5
py = pseq + 4;
*len = tlen;
rewind(fp);
10
while (fgets(line, 1024, fp)) {
    if (*line == ';' || *line == '<' || *line == '>')
        continue;
    for (px = line; *px != '\n'; px++) {
        15
        if (isupper(*px))
            *py++ = *px;
        else if (islower(*px))
            *py++ = toupper(*px);
        if (index("ATGCU", *(py-1)))
            natgc++;
    }
    *py++ = '\0';
    *py = '\0';
20
(void) fclose(fp);
    dna = natgc > (tlen/3);
    return(pseq+4);
}
25
char *
g_calloc(msg, nx, sz)
30
    char *msg;           /* program, calling routine */
    int nx, sz;          /* number and size of elements */
{
    char *px, *calloc();
35
    if ((px = calloc((unsigned)nx, (unsigned)sz)) == 0) {
        if (*msg) {
            fprintf(stderr, "%s: g_calloc() failed %s (n=%d, sz=%d)\n", prog, msg, nx, sz);
            exit(1);
        }
        40
        return(px);
    }
45
/*
 * get final jmps from dx[] or tmp file, set pp[], reset dmax: main()
 */
readjmps()
50
{
    int fd = -1;
    int siz, i0, i1;
    register i, j, xx;
55
    if (fj) {
        (void) fclose(fj);
        if ((fd = open(jname, O_RDONLY, 0)) < 0) {
            fprintf(stderr, "%s: can't open() %s\n", prog, jname);
            cleanup(1);
        }
        60
        for (i = i0 = i1 = 0, dmax0 = dmax, xx = len0; ; i++) {
            while (1) {
                for (j = dx[dmax].jmp; j >= 0 && dx[dmax].jp.x[j] >= xx; j--)
                    ;
            }
        }
    }
}

```

Table 1 (cont')

...readjmps

```
if (j < 0 && dx[dmax].offset && fj) {
    (void) lseek(fd, dx[dmax].offset, 0);
    (void) read(fd, (char *)&dx[dmax].jp, sizeof(struct jmp));
    (void) read(fd, (char *)&dx[dmax].offset, sizeof(dx[dmax].offset));
    dx[dmax].ijmp = MAXJMP-1;
}
else
    break;
}
if (i >= JMPS) {
    fprintf(stderr, "%s: too many gaps in alignment\n", prog);
    cleanup(1);
}
if (j >= 0) {
    siz = dx[dmax].jp.n[j];
    xx = dx[dmax].jp.x[j];
    dmax += siz;
    if (siz < 0) { /* gap in second seq */
        pp[1].n[i1] = -siz;
        xx += siz;
        /* id = xx - yy + len1 - 1
        */
        pp[1].x[i1] = xx - dmax + len1 - 1;
        gapy++;
        ngapy -= siz;
    /* ignore MAXGAP when doing endgaps */
        siz = (-siz < MAXGAP || endgaps)? -siz : MAXGAP;
        i1++;
    }
    else if (siz > 0) { /* gap in first seq */
        pp[0].n[i0] = siz;
        pp[0].x[i0] = xx;
        gapx++;
        ngapx += siz;
    /* ignore MAXGAP when doing endgaps */
        siz = (siz < MAXGAP || endgaps)? siz : MAXGAP;
        i0++;
    }
}
else
    break;
}
/* reverse the order of jmps
 */
for (j = 0, i0--; j < i0; j++, i0--) {
    i = pp[0].n[j]; pp[0].n[j] = pp[0].n[i0]; pp[0].n[i0] = i;
    i = pp[0].x[j]; pp[0].x[j] = pp[0].x[i0]; pp[0].x[i0] = i;
}
for (j = 0, i1--; j < i1; j++, i1--) {
    i = pp[1].n[j]; pp[1].n[j] = pp[1].n[i1]; pp[1].n[i1] = i;
    i = pp[1].x[j]; pp[1].x[j] = pp[1].x[i1]; pp[1].x[i1] = i;
}
if (fd >= 0)
    (void) close(fd);
if (fj) {
    (void) unlink(jname);
    fj = 0;
    offset = 0;
}
}
```

Table 1 (cont')

```
/*
 * write a filled jmp struct offset of the prev one (if any): nw()
 */
5  writejmps(ix)
    int      ix;
{
    char    *mktemp();

10  if (!fj) {
        if (mktemp(jname) < 0) {
            fprintf(stderr, "%s: can't mktemp() %s\n", prog, jname);
            cleanup(1);
        }
15  if ((fj = fopen(jname, "w")) == 0) {
            fprintf(stderr, "%s: can't write %s\n", prog, jname);
            exit(1);
        }
20  (void) fwrite((char *)&dx[ix].jp, sizeof(struct jmp), 1, fj);
    (void) fwrite((char *)&dx[ix].offset, sizeof(dx[ix].offset), 1, fj);
}
25
30
35
40

45

50

55

60
```

Table 2

PRO	XXXXXXXXXXXXXXXXXX	(Length = 15 amino acids)
Comparison Protein	XXXXXYYYYYYY	(Length = 12 amino acids)

5 % amino acid sequence identity =

(the number of identically matching amino acid residues between the two polypeptide sequences as determined by ALIGN-2) divided by (the total number of amino acid residues of the PRO polypeptide) =

10 5 divided by 15 = 33.3%

Copyright © The McGraw-Hill Companies, Inc.

Table 3

PRO	XXXXXXXXXX	(Length = 10 amino acids)
Comparison Protein	XXXXXXXXYYYYYYZZYZ	(Length = 15 amino acids)

5 % amino acid sequence identity =

(the number of identically matching amino acid residues between the two polypeptide sequences as determined by ALIGN-2) divided by (the total number of amino acid residues of the PRO polypeptide) =

10 5 divided by 10 = 50%

Table 4

PRO-DNA	NNNNNNNNNNNNNN	(Length = 14 nucleotides)
Comparison DNA	NNNNNNLLLLLLLL	(Length = 16 nucleotides)

5 % nucleic acid sequence identity =

(the number of identically matching nucleotides between the two nucleic acid sequences as determined by ALIGN-2) divided by (the total number of nucleotides of the PRO-DNA nucleic acid sequence) =

$$10 \quad 6 \text{ divided by } 14 = 42.9\%$$

Table 5

PRO-DNA	NNNNNNNNNNNN	(Length = 12 nucleotides)
Comparison DNA	NNNNLLLVV	(Length = 9 nucleotides)

5 % nucleic acid sequence identity =

(the number of identically matching nucleotides between the two nucleic acid sequences as determined by ALIGN-2) divided by (the total number of nucleotides of the PRO-DNA nucleic acid sequence) =

10 4 divided by 12 = 33.3%

PRO-DNA sequence: NNNNNNNNNNNN

II. Compositions and Methods of the Invention

A. Full-Length PRO Polypeptides

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO polypeptides. In particular, cDNAs encoding various PRO polypeptides have been identified and isolated, as disclosed in further detail in the Examples below. It is noted 5 that proteins produced in separate expression rounds may be given different PRO numbers but the UNQ number is unique for any given DNA and the encoded protein, and will not be changed. However, for sake of simplicity, in the present specification the protein encoded by the full length native nucleic acid molecules disclosed herein as well as all further native homologues and variants included in the foregoing definition of PRO, will be referred to as "PRO/number", regardless of their origin or mode of preparation.

10 As disclosed in the Examples below, various cDNA clones have been deposited with the ATCC. The actual nucleotide sequences of those clones can readily be determined by the skilled artisan by sequencing of the deposited clone using routine methods in the art. The predicted amino acid sequence can be determined from the nucleotide sequence using routine skill. For the PRO polypeptides and encoding nucleic acids described herein, Applicants have identified what is believed to be the reading frame best identifiable with the sequence information available at the time.

1. Full-length PRO241 Polypeptides

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO241. In particular, Applicants have identified and isolated cDNA encoding a PRO241 polypeptide, as disclosed in further detail in the Examples below. Using BLAST and FastA sequence alignment computer programs, Applicants found that portions of the PRO241 polypeptide have certain homology with the various biglycan proteins. Accordingly, it is presently believed that PRO241 polypeptide disclosed in the present application is a newly identified biglycan homolog polypeptide and may possess activity typical of biglycan proteins.

25
20
15
10
5
The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO243. In particular, Applicants have identified and isolated cDNA encoding a PRO243 polypeptide, as disclosed in further detail in the Examples below. Using BLAST, BLAST-2 and FastA sequence alignment computer programs, Applicants found that a full-length native sequence PRO243 (shown in Figure 4 and SEQ ID NO:7) has certain amino acid sequence identity with African clawed frog and Xenopus chordin and certain homology with rat chordin. Accordingly, it is presently believed that PRO243 disclosed in the present application is a newly identified member of the chordin protein family and may possess ability to influence notochord and muscle formation by the dorsalization of the mesoderm.

3. **Full-length PRO299**

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO299. In particular, Applicants have identified and isolated cDNA encoding a PRO299 polypeptide, as disclosed in further detail in the Examples below. Using BLAST and FastA sequence alignment computer programs, Applicants found that various portions of the PRO299 polypeptide have 5 certain homology with the notch protein. Accordingly, it is presently believed that PRO299 polypeptide disclosed in the present application is a newly identified member of the notch protein family and possesses signaling properties typical of the notch protein family.

4. **Full-length PRO323 Polypeptides**

10 The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO323. In particular, Applicants have identified and isolated cDNA encoding a PRO323 polypeptide, as disclosed in further detail in the Examples below. Using BLAST and FastA sequence alignment computer programs, Applicants found that various portions of the PRO323 polypeptide have certain homology with various dipeptidase proteins. Accordingly, it is presently believed that PRO323 polypeptide disclosed in the present application is a newly identified dipeptidase homolog that has dipeptidase activity

5. **Full-length PRO327 Polypeptides**

20 The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO327. In particular, Applicants have identified and isolated cDNA encoding a PRO327 polypeptide, as disclosed in further detail in the Examples below. Using BLAST and FastA sequence alignment computer programs, Applicants found that portions of the PRO327 polypeptide have certain homology with various prolactin receptor proteins. Accordingly, it is presently believed that PRO327 polypeptide disclosed in the present application is a newly identified prolactin receptor homolog and has activity 25 typical of a prolactin receptor protein.

6. **Full-length PRO233 Polypeptides**

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO233. In particular, Applicants have identified and isolated cDNA 30 encoding a PRO233 polypeptide, as disclosed in further detail in the Examples below. Using BLAST and FastA sequence alignment computer programs, Applicants found that various portions of the PRO233 polypeptide have certain homology with various reductase proteins. Applicants have also found that the DNA encoding the PRO233 polypeptide has significant homology with proteins from *Caenorhabditis elegans*. Accordingly, it is presently believed that PRO233 polypeptide disclosed in the present application is a newly identified member 35 of the reductase family and possesses the ability to effect the redox state of a cell typical of the reductase family.

7. **Full-length PRO344 Polypeptides**

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO344. In particular, Applicants have identified and isolated cDNA encoding PRO344 polypeptides, as disclosed in further detail in the Examples below. Using BLAST and FastA sequence alignment computer programs, Applicants found that various portions of the PRO344 polypeptide have 5 certain homology with the human and mouse complement proteins. Accordingly, it is presently believed that the PRO344 polypeptide disclosed in the present application is a newly identified member of the complement family and possesses the ability to affect the inflammation process as is typical of the complement family of proteins.

10 8. **Full-length PRO347 Polypeptides**

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO347. In particular, Applicants have identified and isolated cDNA encoding a PRO347 polypeptide, as disclosed in further detail in the Examples below. Using BLAST and FastA sequence alignment computer programs, Applicants found that portions of the PRO347 polypeptide have certain homology with various cysteine-rich secretory proteins. Accordingly, it is presently believed that PRO347 polypeptide disclosed in the present application is a newly identified cysteine-rich secretory protein and may possess activity typical of the cysteine-rich secretory protein family.

15 9. **Full-length PRO354 Polypeptides**

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO354. In particular, Applicants have identified and isolated cDNA encoding a PRO354 polypeptide, as disclosed in further detail in the Examples below. Using BLAST and FastA sequence alignment computer programs, Applicants found that portions of the PRO354 polypeptide have certain homology with the inter-alpha-trypsin inhibitor heavy chain protein. Accordingly, it is presently believed that 25 PRO354 polypeptide disclosed in the present application is a newly identified inter-alpha-trypsin inhibitor heavy chain homolog.

20 10. **Full-length PRO355 Polypeptides**

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO355. In particular, Applicants have identified and isolated cDNA encoding a PRO355 polypeptide, as disclosed in further detail in the Examples below. Using BLAST and FastA sequence alignment computer programs, Applicants found that various portions of the PRO355 polypeptide have certain homology with the CRTAM protein. Applicants have also found that the DNA encoding the PRO355 polypeptide also has homology to the thymocyte activation and developmental protein, the H20A receptor, the 35 H20B receptor, the poliovirus receptor and the *Cercopithecus aethiops* AGM delta 1 protein. Accordingly, it is presently believed that PRO355 polypeptide disclosed in the present application is a newly identified member of the CRTAM protein family.

11. **Full-length PRO357 Polypeptides**

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO357. In particular, Applicants have identified and isolated cDNA encoding a PRO357 polypeptide, as disclosed in further detail in the Examples below. Using BLAST and FastA sequence alignment computer programs, Applicants found that various portions of the PRO357 polypeptide have 5 certain homology with the acid labile subunit of insulin-like growth factor. Applicants have also found that non-coding regions of the DNA44804-1248 align with a human gene signature as described in WO 95/14772. Applicants have further found that non-coding regions of the DNA44804-1248 align with the adenovirus type 10 12/human recombinant viral DNA as described in Deuring and Doerfler, *Gene*, 26:283-289 (1983). Based on the coding region homology, it is presently believed that PRO357 polypeptide disclosed in the present application 15 is a newly identified member of the leucine rich repeat family of proteins, and particularly, is related to the acid labile subunit of insulin-like growth factor. As such, PRO357 is likely to be involved in binding mechanisms, and may be part of a complex.

12. **Full-length PRO715 Polypeptides**

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO715. In particular, Applicants have identified and isolated cDNA molecules encoding PRO715 polypeptides, as disclosed in further detail in the Examples below. Using BLAST and FastA sequence alignment computer programs, Applicants found that various portions of the PRO715 polypeptides have certain homology with the various members of the tumor necrosis family of proteins. Accordingly, it is presently believed that the PRO715 polypeptides disclosed in the present application are newly 20 identified members of the tumor necrosis factor family of proteins.

13. **Full-length PRO353 Polypeptides**

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO353. In particular, Applicants have identified and isolated cDNA 25 encoding PRO353 polypeptides, as disclosed in further detail in the Examples below. Using BLAST and, FastA sequence alignment computer programs, Applicants found that various portions of the PRO353 polypeptides have certain homology with the human and mouse complement proteins. Accordingly, it is presently believed that the PRO353 polypeptides disclosed in the present application are newly identified members of the complement 30 protein family and possesses the ability to effect the inflammation process as is typical of the complement family of proteins.

14. **Full-length PRO361 Polypeptides**

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO361. In particular, Applicants have identified and isolated cDNA 35 encoding a PRO361 polypeptide, as disclosed in further detail in the Examples below. Using BLAST and FastA sequence alignment computer programs, Applicants found that various portions of the PRO361 polypeptide have

certain homology with the mucin and chitinase proteins. Accordingly, it is presently believed that PRO361 polypeptide disclosed in the present application is a newly identified member of the mucin and/or chitinase protein families and may be associated with cancer, plant pathogenesis or receptor functions typical of the mucin and chitinase protein families, respectively.

5

15. Full-length PRO365 Polypeptides

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO365. In particular, Applicants have identified and isolated cDNA encoding a PRO365 polypeptide, as disclosed in further detail in the Examples below. Using BLAST and FastA sequence alignment computer programs, Applicants found that various portions of the PRO365 polypeptide have 10 certain homology with the human 2-19 protein. Accordingly, it is presently believed that PRO365 polypeptide disclosed in the present application is a newly identified member of the human 2-19 protein family.

2. PRO Polypeptide Variants

15

In addition to the full-length native sequence PRO polypeptides described herein, it is contemplated that PRO variants can be prepared. PRO variants can be prepared by introducing appropriate nucleotide changes into the PRO DNA, and/or by synthesis of the desired PRO polypeptide. Those skilled in the art will appreciate that amino acid changes may alter post-translational processes of the PRO, such as changing the number or position of glycosylation sites or altering the membrane anchoring characteristics.

20
25

Variations in the native full-length sequence PRO or in various domains of the PRO described herein, can be made, for example, using any of the techniques and guidelines for conservative and non-conservative mutations set forth, for instance, in U.S. Patent No. 5,364,934. Variations may be a substitution, deletion or insertion of one or more codons encoding the PRO that results in a change in the amino acid sequence of the PRO as compared with the native sequence PRO. Optionally the variation is by substitution of at least one amino acid with any other amino acid in one or more of the domains of the PRO. Guidance in determining which amino acid residue may be inserted, substituted or deleted without adversely affecting the desired activity may be found by comparing the sequence of the PRO with that of homologous known protein molecules and minimizing the number of amino acid sequence changes made in regions of high homology. Amino acid substitutions can be the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, such as the replacement of a leucine with a serine, i.e., conservative amino acid 30 replacements. Insertions or deletions may optionally be in the range of about 1 to 5 amino acids. The variation allowed may be determined by systematically making insertions, deletions or substitutions of amino acids in the sequence and testing the resulting variants for activity exhibited by the full-length or mature native sequence.

35

PRO polypeptide fragments are provided herein. Such fragments may be truncated at the N-terminus or C-terminus, or may lack internal residues, for example, when compared with a full length native protein. Certain fragments lack amino acid residues that are not essential for a desired biological activity of the PRO polypeptide.

PRO fragments may be prepared by any of a number of conventional techniques. Desired peptide fragments may be chemically synthesized. An alternative approach involves generating PRO fragments by enzymatic digestion, e.g., by treating the protein with an enzyme known to cleave proteins at sites defined by particular amino acid residues, or by digesting the DNA with suitable restriction enzymes and isolating the desired fragment. Yet another suitable technique involves isolating and amplifying a DNA fragment encoding 5 a desired polypeptide fragment, by polymerase chain reaction (PCR). Oligonucleotides that define the desired termini of the DNA fragment are employed at the 5' and 3' primers in the PCR. Preferably, PRO polypeptide fragments share at least one biological and/or immunological activity with the native PRO polypeptide disclosed herein.

In particular embodiments, conservative substitutions of interest are shown in Table 1 under the heading 10 of preferred substitutions. If such substitutions result in a change in biological activity, then more substantial changes, denominated exemplary substitutions in Table 6, or as further described below in reference to amino acid classes, are introduced and the products screened.

Table 6

<u>Original Residue</u>	<u>Exemplary Substitutions</u>	<u>Preferred Substitutions</u>
5 Ala (A)	val; leu; ile	val
Arg (R)	lys; gln; asn	lys
Asn (N)	gln; his; lys; arg	gln
Asp (D)	glu	glu
Cys (C)	ser	ser
10 Gln (Q)	asn	asn
Glu (E)	asp	asp
Gly (G)	pro; ala	ala
His (H)	asn; gln; lys; arg	arg
Ile (I)	leu; val; met; ala; phe; norleucine	leu
15 Leu (L)	norleucine; ile; val; met; ala; phe	ile
Lys (K)	arg; gln; asn	arg
Met (M)	leu; phe; ile	leu
20 Phe (F)	leu; val; ile; ala; tyr	leu
Pro (P)	ala	ala
Ser (S)	thr	thr
Thr (T)	ser	ser
Trp (W)	tyr; phe	tyr
25 Tyr (Y)	trp; phe; thr; ser	phe
Val (V)	ile; leu; met; phe; ala; norleucine	leu

Substantial modifications in function or immunological identity of the PRO polypeptide are accomplished by selecting substitutions that differ significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. Naturally occurring residues are divided into groups based on common side-chain properties:

- (1) hydrophobic: norleucine, met, ala, val, leu, ile;
- 35 (2) neutral hydrophilic: cys, ser, thr;
- (3) acidic: asp, glu;
- (4) basic: asn, gln, his, lys, arg;
- (5) residues that influence chain orientation: gly, pro; and
- (6) aromatic: trp, tyr, phe.

40 Non-conservative substitutions will entail exchanging a member of one of these classes for another class. Such substituted residues also may be introduced into the conservative substitution sites or, more preferably, into the remaining (non-conserved) sites.

The variations can be made using methods known in the art such as oligonucleotide-mediated (site-directed) mutagenesis, alanine scanning, and PCR mutagenesis. Site-directed mutagenesis [Carter et al., Nucl. Acids Res., 13:4331 (1986); Zoller et al., Nucl. Acids Res., 10:6487 (1987)], cassette mutagenesis [Wells et al., Gene, 34:315 (1985)], restriction selection mutagenesis [Wells et al., Philos. Trans. R. Soc. London Ser. A, 317:415 (1986)] or other known techniques can be performed on the cloned DNA to produce the PRO variant

DNA.

Scanning amino acid analysis can also be employed to identify one or more amino acids along a contiguous sequence. Among the preferred scanning amino acids are relatively small, neutral amino acids. Such amino acids include alanine, glycine, serine, and cysteine. Alanine is typically a preferred scanning amino acid among this group because it eliminates the side-chain beyond the beta-carbon and is less likely to alter the main-chain conformation of the variant [Cunningham and Wells, *Science*, 244: 1081-1085 (1989)]. Alanine is also typically preferred because it is the most common amino acid. Further, it is frequently found in both buried and exposed positions [Creighton, *The Proteins*, (W.H. Freeman & Co., N.Y.); Chothia, *J. Mol. Biol.*, 150:1 (1976)]. If alanine substitution does not yield adequate amounts of variant, an isoteric amino acid can be used.

10 C. Modifications of PRO

Covalent modifications of PRO are included within the scope of this invention. One type of covalent modification includes reacting targeted amino acid residues of a PRO polypeptide with an organic derivatizing agent that is capable of reacting with selected side chains or the N- or C-terminal residues of the PRO. Derivatization with bifunctional agents is useful, for instance, for crosslinking PRO to a water-insoluble support matrix or surface for use in the method for purifying anti-PRO antibodies, and vice-versa. Commonly used crosslinking agents include, e.g., 1,1-bis(diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), bifunctional maleimides such as bis-N-maleimido-1,8-octane and agents such as methyl-3-[(p-azidophenyl)dithio]propioimidate.

Other modifications include deamidation of glutaminyl and asparaginyl residues to the corresponding glutamyl and aspartyl residues, respectively, hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the α -amino groups of lysine, arginine, and histidine side chains [T.E. Creighton, *Proteins: Structure and Molecular Properties*, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)], acetylation of the N-terminal amine, and amidation of any C-terminal carboxyl group.

25 Another type of covalent modification of the PRO polypeptide included within the scope of this invention comprises altering the native glycosylation pattern of the polypeptide. "Altering the native glycosylation pattern" is intended for purposes herein to mean deleting one or more carbohydrate moieties found in native sequence PRO (either by removing the underlying glycosylation site or by deleting the glycosylation by chemical and/or enzymatic means), and/or adding one or more glycosylation sites that are not present in the 30 native sequence PRO. In addition, the phrase includes qualitative changes in the glycosylation of the native proteins, involving a change in the nature and proportions of the various carbohydrate moieties present.

35 Addition of glycosylation sites to the PRO polypeptide may be accomplished by altering the amino acid sequence. The alteration may be made, for example, by the addition of, or substitution by, one or more serine or threonine residues to the native sequence PRO (for O-linked glycosylation sites). The PRO amino acid sequence may optionally be altered through changes at the DNA level, particularly by mutating the DNA encoding the PRO polypeptide at preselected bases such that codons are generated that will translate into the desired amino acids.

Another means of increasing the number of carbohydrate moieties on the PRO polypeptide is by chemical or enzymatic coupling of glycosides to the polypeptide. Such methods are described in the art, e.g., in WO 87/05330 published 11 September 1987, and in Aplin and Wriston, CRC Crit. Rev. Biochem., pp. 259-306 (1981).

Removal of carbohydrate moieties present on the PRO polypeptide may be accomplished chemically or enzymatically or by mutational substitution of codons encoding for amino acid residues that serve as targets for glycosylation. Chemical deglycosylation techniques are known in the art and described, for instance, by Hakimuddin, et al., Arch. Biochem. Biophys., 259:52 (1987) and by Edge et al., Anal. Biochem., 118:131 (1981). Enzymatic cleavage of carbohydrate moieties on polypeptides can be achieved by the use of a variety of endo- and exo-glycosidases as described by Thotakura et al., Meth. Enzymol., 138:350 (1987).

Another type of covalent modification of PRO comprises linking the PRO polypeptide to one of a variety of nonproteinaceous polymers, e.g., polyethylene glycol (PEG), polypropylene glycol, or polyoxyalkylenes, in the manner set forth in U.S. Patent Nos. 4,640,835; 4,496,689; 4,301,144; 4,670,417; 4,791,192 or 4,179,337.

The PRO of the present invention may also be modified in a way to form a chimeric molecule comprising PRO fused to another, heterologous polypeptide or amino acid sequence.

In one embodiment, such a chimeric molecule comprises a fusion of the PRO with a tag polypeptide which provides an epitope to which an anti-tag antibody can selectively bind. The epitope tag is generally placed at the amino- or carboxyl- terminus of the PRO. The presence of such epitope-tagged forms of the PRO can be detected using an antibody against the tag polypeptide. Also, provision of the epitope tag enables the PRO to be readily purified by affinity purification using an anti-tag antibody or another type of affinity matrix that binds to the epitope tag. Various tag polypeptides and their respective antibodies are well known in the art. Examples include poly-histidine (poly-his) or poly-histidine-glycine (poly-his-gly) tags; the flu HA tag polypeptide and its antibody 12CA5 [Field et al., Mol. Cell. Biol., 8:2159-2165 (1988)]; the c-myc tag and the 8F9, 3C7, 6E10, G4, B7 and 9E10 antibodies thereto [Evan et al., Molecular and Cellular Biology, 5:3610-3616 (1985)]; and the Herpes Simplex virus glycoprotein D (gD) tag and its antibody [Paborsky et al., Protein Engineering, 3(6):547-553 (1990)]. Other tag polypeptides include the Flag-peptide [Hopp et al., BioTechnology, 6:1204-1210 (1988)]; the KT3 epitope peptide [Martin et al., Science, 255:192-194 (1992)]; an α -tubulin epitope peptide [Skinner et al., J. Biol. Chem., 266:15163-15166 (1991)]; and the T7 gene 10 protein peptide tag [Lutz-Freyermuth et al., Proc. Natl. Acad. Sci. USA, 87:6393-6397 (1990)].

In an alternative embodiment, the chimeric molecule may comprise a fusion of the PRO with an immunoglobulin or a particular region of an immunoglobulin. For a bivalent form of the chimeric molecule (also referred to as an "immunoadhesin"), such a fusion could be to the Fc region of an IgG molecule. The Ig fusions preferably include the substitution of a soluble (transmembrane domain deleted or inactivated) form of a PRO polypeptide in place of at least one variable region within an Ig molecule. In a particularly preferred embodiment, the immunoglobulin fusion includes the hinge, CH2 and CH3, or the hinge, CH1, CH2 and CH3 regions of an IgG1 molecule. For the production of immunoglobulin fusions see also US Patent No. 5,428,130 issued June 27, 1995.

D. Preparation of PRO

The description below relates primarily to production of PRO by culturing cells transformed or transfected with a vector containing PRO nucleic acid. It is, of course, contemplated that alternative methods, which are well known in the art, may be employed to prepare PRO. For instance, the PRO sequence, or portions thereof, may be produced by direct peptide synthesis using solid-phase techniques [see, e.g., Stewart 5 et al., Solid-Phase Peptide Synthesis, W.H. Freeman Co., San Francisco, CA (1969); Merrifield, J. Am. Chem. Soc., 85:2149-2154 (1963)]. *In vitro* protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be accomplished, for instance, using an Applied Biosystems Peptide Synthesizer (Foster City, CA) using manufacturer's instructions. Various portions of the PRO may be chemically synthesized separately and combined using chemical or enzymatic methods to produce the full-length 10 PRO.

1. Isolation of DNA Encoding PRO

DNA encoding PRO may be obtained from a cDNA library prepared from tissue believed to possess the PRO mRNA and to express it at a detectable level. Accordingly, human PRO DNA can be conveniently obtained from a cDNA library prepared from human tissue, such as described in the Examples. The PRO-encoding gene may also be obtained from a genomic library or by known synthetic procedures (e.g., automated nucleic acid synthesis).

Libraries can be screened with probes (such as antibodies to the PRO or oligonucleotides of at least about 20-80 bases) designed to identify the gene of interest or the protein encoded by it. Screening the cDNA or genomic library with the selected probe may be conducted using standard procedures, such as described in Sambrook et al., Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Laboratory Press, 1989). An alternative means to isolate the gene encoding PRO is to use PCR methodology [Sambrook et al., supra; Dieffenbach et al., PCR Primer: A Laboratory Manual (Cold Spring Harbor Laboratory Press, 1995)].

The Examples below describe techniques for screening a cDNA library. The oligonucleotide sequences 25 selected as probes should be of sufficient length and sufficiently unambiguous that false positives are minimized. The oligonucleotide is preferably labeled such that it can be detected upon hybridization to DNA in the library being screened. Methods of labeling are well known in the art, and include the use of radiolabels like ^{32}P -labeled ATP, biotinylation or enzyme labeling. Hybridization conditions, including moderate stringency and high stringency, are provided in Sambrook et al., supra.

30 Sequences identified in such library screening methods can be compared and aligned to other known sequences deposited and available in public databases such as GenBank or other private sequence databases. Sequence identity (at either the amino acid or nucleotide level) within defined regions of the molecule or across the full-length sequence can be determined using methods known in the art and as described herein.

Nucleic acid having protein coding sequence may be obtained by screening selected cDNA or genomic 35 libraries using the deduced amino acid sequence disclosed herein for the first time, and, if necessary, using conventional primer extension procedures as described in Sambrook et al., supra, to detect precursors and processing intermediates of mRNA that may not have been reverse-transcribed into cDNA.

2. Selection and Transformation of Host Cells

Host cells are transfected or transformed with expression or cloning vectors described herein for PRO production and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. The culture conditions, such as media, temperature, pH and the like, can be selected by the skilled artisan without undue experimentation. In general, 5 principles, protocols, and practical techniques for maximizing the productivity of cell cultures can be found in Mammalian Cell Biotechnology: a Practical Approach, M. Butler, ed. (IRL Press, 1991) and Sambrook et al., supra.

Methods of eukaryotic cell transfection and prokaryotic cell transformation are known to the ordinarily skilled artisan, for example, CaCl_2 , CaPO_4 , liposome-mediated and electroporation. Depending on the host cell 10 used, transformation is performed using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described in Sambrook et al., supra, or electroporation is generally used for prokaryotes. Infection with *Agrobacterium tumefaciens* is used for transformation of certain plant cells, as described by Shaw et al., Gene, 23:315 (1983) and WO 89/05859 published 29 June 1989. For mammalian cells without such cell walls, the calcium phosphate precipitation method of Graham and van der Eb, Virology, 15 52:456-457 (1978) can be employed. General aspects of mammalian cell host system transfections have been described in U.S. Patent No. 4,399,216. Transformations into yeast are typically carried out according to the method of Van Solingen et al., J. Bact., 130:946 (1977) and Hsiao et al., Proc. Natl. Acad. Sci. (USA), 76:3829 (1979). However, other methods for introducing DNA into cells, such as by nuclear microinjection, 20 electroporation, bacterial protoplast fusion with intact cells, or polycations, e.g., polybrene, polyornithine, may also be used. For various techniques for transforming mammalian cells, see Keown et al., Methods in Enzymology, 185:527-537 (1990) and Mansour et al., Nature, 336:348-352 (1988).

Suitable host cells for cloning or expressing the DNA in the vectors herein include prokaryote, yeast, or higher eukaryote cells. Suitable prokaryotes include but are not limited to eubacteria, such as Gram-negative or Gram-positive organisms, for example, Enterobacteriaceae such as *E. coli*. Various *E. coli* strains are 25 publicly available, such as *E. coli* K12 strain MM294 (ATCC 31,446); *E. coli* X1776 (ATCC 31,537); *E. coli* strain W3110 (ATCC 27,325) and K5 772 (ATCC 53,635). Other suitable prokaryotic host cells include Enterobacteriaceae such as *Escherichia*, e.g., *E. coli*, *Enterobacter*, *Erwinia*, *Klebsiella*, *Proteus*, *Salmonella*, e.g., *Salmonella typhimurium*, *Serratia*, e.g., *Serratia marcescans*, and *Shigella*, as well as *Bacilli* such as *B. subtilis* and *B. licheniformis* (e.g., *B. licheniformis* 41P disclosed in DD 266,710 published 12 April 1989), 30 *Pseudomonas* such as *P. aeruginosa*, and *Streptomyces*. These examples are illustrative rather than limiting. Strain W3110 is one particularly preferred host or parent host because it is a common host strain for recombinant DNA product fermentations. Preferably, the host cell secretes minimal amounts of proteolytic enzymes. For example, strain W3110 may be modified to effect a genetic mutation in the genes encoding proteins endogenous to the host, with examples of such hosts including *E. coli* W3110 strain 1A2, which has the complete genotype 35 *tonA*; *E. coli* W3110 strain 9E4, which has the complete genotype *tonA ptr3*; *E. coli* W3110 strain 27C7 (ATCC 55,244), which has the complete genotype *tonA ptr3 phoA E15 (argF-lac)169 degP ompT kan'*; *E. coli* W3110 strain 37D6, which has the complete genotype *tonA ptr3 phoA E15 (argF-lac)169 degP ompT rbs7*

ilvG kan'; *E. coli* W3110 strain 40B4, which is strain 37D6 with a non-kanamycin resistant *degP* deletion mutation; and an *E. coli* strain having mutant periplasmic protease disclosed in U.S. Patent No. 4,946,783 issued 7 August 1990. Alternatively, *in vitro* methods of cloning, e.g., PCR or other nucleic acid polymerase reactions, are suitable.

In addition to prokaryotes, eukaryotic microbes such as filamentous fungi or yeast are suitable cloning 5 or expression hosts for PRO-encoding vectors. *Saccharomyces cerevisiae* is a commonly used lower eukaryotic host microorganism. Others include *Schizosaccharomyces pombe* (Beach and Nurse, *Nature*, 290: 140 [1981]; EP 139,383 published 2 May 1985); *Kluyveromyces* hosts (U.S. Patent No. 4,943,529; Fleer et al., *Bio/Technology*, 9:968-975 (1991)) such as, e.g., *K. lactis* (MW98-8C, CBS683, CBS4574; Louvencourt et al., *J. Bacteriol.*, 154(2):737-742 [1983]), *K. fragilis* (ATCC 12,424), *K. bulgaricus* (ATCC 16,045), *K. wickeramii* 10 (ATCC 24,178), *K. waltii* (ATCC 56,500), *K. drosophilicola* (ATCC 36,906; Van den Berg et al., *Bio/Technology*, 8:135 (1990)), *K. thermotolerans*, and *K. marxianus*; *yarrowia* (EP 402,226); *Pichia pastoris* (EP 183,070; Sreekrishna et al., *J. Basic Microbiol.*, 28:265-278 [1988]); *Candida*; *Trichoderma reesia* (EP 244,234); *Neurospora crassa* (Case et al., *Proc. Natl. Acad. Sci. USA*, 76:5259-5263 [1979]); *Schwanniomyces* 15 such as *Schwanniomyces occidentalis* (EP 394,538 published 31 October 1990); and filamentous fungi such as, e.g., *Neurospora*, *Penicillium*, *Tolypocladium* (WO 91/00357 published 10 January 1991), and *Aspergillus* hosts 20 such as *A. nidulans* (Ballance et al., *Biochem. Biophys. Res. Commun.*, 112:284-289 [1983]; Tilburn et al., *Gene*, 26:205-221 [1983]; Yelton et al., *Proc. Natl. Acad. Sci. USA*, 81: 1470-1474 [1984]) and *A. niger* (Kelly and Hynes, *EMBO J.*, 4:475-479 [1985]). Methylotropic yeasts are suitable herein and include, but are not limited to, yeast capable of growth on methanol selected from the genera consisting of *Hansenula*, *Candida*, *Kloeckera*, *Pichia*, *Saccharomyces*, *Torulopsis*, and *Rhodotorula*. A list of specific species that are exemplary 25 of this class of yeasts may be found in C. Anthony, *The Biochemistry of Methylotrophs*, 269 (1982).

Suitable host cells for the expression of glycosylated PRO are derived from multicellular organisms. Examples of invertebrate cells include insect cells such as *Drosophila* S2 and *Spodoptera* Sf9, as well as plant 25 cells. Examples of useful mammalian host cell lines include Chinese hamster ovary (CHO) and COS cells. More specific examples include monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham et al., *J. Gen Virol.*, 36:59 (1977)); Chinese hamster ovary cells/-DHFR (CHO, Urlaub and Chasin, *Proc. Natl. Acad. Sci. USA*, 77:4216 (1980)); mouse sertoli cells (TM4, Mather, *Biol. Reprod.*, 23:243-251 (1980)); human lung 30 cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); and mouse mammary tumor (MMT 060562, ATCC CCL51). The selection of the appropriate host cell is deemed to be within the skill in the art.

3. Selection and Use of a Replicable Vector

The nucleic acid (e.g., cDNA or genomic DNA) encoding PRO may be inserted into a replicable vector 35 for cloning (amplification of the DNA) or for expression. Various vectors are publicly available. The vector may, for example, be in the form of a plasmid, cosmid, viral particle, or phage. The appropriate nucleic acid sequence may be inserted into the vector by a variety of procedures. In general, DNA is inserted into an appropriate restriction endonuclease site(s) using techniques known in the art. Vector components generally

include, but are not limited to, one or more of a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence. Construction of suitable vectors containing one or more of these components employs standard ligation techniques which are known to the skilled artisan.

The PRO may be produced recombinantly not only directly, but also as a fusion polypeptide with a 5 heterologous polypeptide, which may be a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature protein or polypeptide. In general, the signal sequence may be a component of the vector, or it may be a part of the PRO-encoding DNA that is inserted into the vector. The signal sequence may be a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leaders. For yeast secretion the signal sequence may be, e.g., 10 the yeast invertase leader, alpha factor leader (including *Saccharomyces* and *Kluyveromyces* α -factor leaders, the latter described in U.S. Patent No. 5,010,182), or acid phosphatase leader, the *C. albicans* glucoamylase leader (EP 362,179 published 4 April 1990), or the signal described in WO 90/13646 published 15 November 1990. In mammalian cell expression, mammalian signal sequences may be used to direct secretion of the 15 protein, such as signal sequences from secreted polypeptides of the same or related species, as well as viral secretory leaders.

Both expression and cloning vectors contain a nucleic acid sequence that enables the vector to replicate in one or more selected host cells. Such sequences are well known for a variety of bacteria, yeast, and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2μ plasmid origin is suitable for yeast, and various viral origins (SV40, polyoma, adenovirus, VSV or BPV) are useful for 20 cloning vectors in mammalian cells.

Expression and cloning vectors will typically contain a selection gene, also termed a selectable marker. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, e.g., ampicillin, neomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or (c) supply critical nutrients not available from complex media, e.g., the gene encoding D-alanine racemase for *Bacilli*.

25 An example of suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up the PRO-encoding nucleic acid, such as DHFR or thymidine kinase. An appropriate host cell when wild-type DHFR is employed is the CHO cell line deficient in DHFR activity, prepared and propagated as described by Urlaub et al., Proc. Natl. Acad. Sci. USA, 77:4216 (1980). A suitable selection gene for use in yeast is the *trp1* gene present in the yeast plasmid YRp7 [Stinchcomb et al., Nature, 30 282:39 (1979); Kingsman et al., Gene, 7:141 (1979); Tschemper et al., Gene, 10:157 (1980)]. The *trp1* gene provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example, ATCC No. 44076 or PEP4-1 [Jones, Genetics, 85:12 (1977)].

35 Expression and cloning vectors usually contain a promoter operably linked to the PRO-encoding nucleic acid sequence to direct mRNA synthesis. Promoters recognized by a variety of potential host cells are well known. Promoters suitable for use with prokaryotic hosts include the β -lactamase and lactose promoter systems [Chang et al., Nature, 275:615 (1978); Goeddel et al., Nature, 281:544 (1979)], alkaline phosphatase, a tryptophan (*trp*) promoter system [Goeddel, Nucleic Acids Res., 8:4057 (1980); EP 36,776], and hybrid

promoters such as the tac promoter [deBoer et al., Proc. Natl. Acad. Sci. USA, 80:21-25 (1983)]. Promoters for use in bacterial systems also will contain a Shine-Dalgarno (S.D.) sequence operably linked to the DNA encoding PRO.

Examples of suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglycerate kinase [Hitzeman et al., J. Biol. Chem., 255:2073 (1980)] or other glycolytic enzymes [Hess 5 et al., J. Adv. Enzyme Reg., 7:149 (1968); Holland, Biochemistry, 17:4900 (1978)], such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other yeast promoters, which are inducible promoters having the additional advantage of transcription 10 controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytchrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in EP 73,657.

PRO transcription from vectors in mammalian host cells is controlled, for example, by promoters obtained from the genomes of viruses such as polyoma virus, fowlpox virus (UK 2,211,504 published 5 July 1989), adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and Simian Virus 40 (SV40), from heterologous mammalian promoters, e.g., the actin promoter or an immunoglobulin promoter, and from heat-shock promoters, provided such promoters are compatible with the host cell systems.

Transcription of a DNA encoding the PRO by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp, that act on a promoter to increase its transcription. Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, α -fetoprotein, and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication 25 origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. The enhancer may be spliced into the vector at a position 5' or 3' to the PRO coding sequence, but is preferably located at a site 5' from the promoter.

Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human, or nucleated cells from other multicellular organisms) will also contain sequences necessary for the termination of transcription and for stabilizing the mRNA. Such sequences are commonly available from the 5' and, occasionally 3', untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding PRO.

Still other methods, vectors, and host cells suitable for adaptation to the synthesis of PRO in recombinant vertebrate cell culture are described in Gething et al., Nature, 293:620-625 (1981); Mantei et al., 35 Nature, 281:40-46 (1979); EP 117,060; and EP 117,058.

4. Detecting Gene Amplification/Expression

Gene amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA [Thomas, Proc. Natl. Acad. Sci. USA, 77:5201-5205 (1980)], dot blotting (DNA analysis), or *in situ* hybridization, using an appropriately labeled probe, based on the sequences provided herein. Alternatively, antibodies may be employed 5 that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes. The antibodies in turn may be labeled and the assay may be carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected.

Gene expression, alternatively, may be measured by immunological methods, such as 10 immunohistochemical staining of cells or tissue sections and assay of cell culture or body fluids, to quantitate directly the expression of gene product. Antibodies useful for immunohistochemical staining and/or assay of sample fluids may be either monoclonal or polyclonal, and may be prepared in any mammal. Conveniently, the antibodies may be prepared against a native sequence PRO polypeptide or against a synthetic peptide based on the DNA sequences provided herein or against exogenous sequence fused to PRO DNA and encoding a specific 15 antibody epitope.

5. Purification of Polypeptide

Forms of PRO may be recovered from culture medium or from host cell lysates. If membrane-bound, it can be released from the membrane using a suitable detergent solution (e.g. Triton-X 100) or by enzymatic 20 cleavage. Cells employed in expression of PRO can be disrupted by various physical or chemical means, such as freeze-thaw cycling, sonication, mechanical disruption, or cell lysing agents.

It may be desired to purify PRO from recombinant cell proteins or polypeptides. The following 25 procedures are exemplary of suitable purification procedures: by fractionation on an ion-exchange column; ethanol precipitation; reverse phase HPLC; chromatography on silica or on a cation-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; gel filtration using, for example, 30 Sephadex G-75; protein A Sepharose columns to remove contaminants such as IgG; and metal chelating columns to bind epitope-tagged forms of the PRO. Various methods of protein purification may be employed and such methods are known in the art and described for example in Deutscher, Methods in Enzymology, 182 (1990); Scopes, Protein Purification: Principles and Practice, Springer-Verlag, New York (1982). The purification step(s) selected will depend, for example, on the nature of the production process used and the particular PRO produced.

E. Uses for PRO

Nucleotide sequences (or their complement) encoding PRO have various applications in the art of 35 molecular biology, including uses as hybridization probes, in chromosome and gene mapping and in the generation of anti-sense RNA and DNA. PRO nucleic acid will also be useful for the preparation of PRO polypeptides by the recombinant techniques described herein.

The full-length native sequence PRO gene, or portions thereof, may be used as hybridization probes for a cDNA library to isolate the full-length PRO cDNA or to isolate still other cDNAs (for instance, those encoding naturally-occurring variants of PRO or PRO from other species) which have a desired sequence identity to the native PRO sequence disclosed herein. Optionally, the length of the probes will be about 20 to about 50 bases. The hybridization probes may be derived from at least partially novel regions of the full length native 5 nucleotide sequence wherein those regions may be determined without undue experimentation or from genomic sequences including promoters, enhancer elements and introns of native sequence PRO. By way of example, a screening method will comprise isolating the coding region of the PRO gene using the known DNA sequence to synthesize a selected probe of about 40 bases. Hybridization probes may be labeled by a variety of labels, including radionucleotides such as ^{32}P or ^{35}S , or enzymatic labels such as alkaline phosphatase coupled to the 10 probe via avidin/biotin coupling systems. Labeled probes having a sequence complementary to that of the PRO gene of the present invention can be used to screen libraries of human cDNA, genomic DNA or mRNA to determine which members of such libraries the probe hybridizes to. Hybridization techniques are described in further detail in the Examples below.

Any EST sequences disclosed in the present application may similarly be employed as probes, using the methods disclosed herein.

Other useful fragments of the PRO nucleic acids include antisense or sense oligonucleotides comprising a single-stranded nucleic acid sequence (either RNA or DNA) capable of binding to target PRO mRNA (sense) or PRO DNA (antisense) sequences. Antisense or sense oligonucleotides, according to the present invention, comprise a fragment of the coding region of PRO DNA. Such a fragment generally comprises at least about 14 15 nucleotides, preferably from about 14 to 30 nucleotides. The ability to derive an antisense or a sense 20 oligonucleotide, based upon a cDNA sequence encoding a given protein is described in, for example, Stein and Cohen (Cancer Res. 48:2659, 1988) and van der Krol et al. (BioTechniques 6:958, 1988).

Binding of antisense or sense oligonucleotides to target nucleic acid sequences results in the formation of duplexes that block transcription or translation of the target sequence by one of several means, including 25 enhanced degradation of the duplexes, premature termination of transcription or translation, or by other means. The antisense oligonucleotides thus may be used to block expression of PRO proteins. Antisense or sense oligonucleotides further comprise oligonucleotides having modified sugar-phosphodiester backbones (or other sugar linkages, such as those described in WO 91/06629) and wherein such sugar linkages are resistant to endogenous nucleases. Such oligonucleotides with resistant sugar linkages are stable *in vivo* (i.e., capable of 30 resisting enzymatic degradation) but retain sequence specificity to be able to bind to target nucleotide sequences.

Other examples of sense or antisense oligonucleotides include those oligonucleotides which are covalently linked to organic moieties, such as those described in WO 90/10048, and other moieties that increases 35 affinity of the oligonucleotide for a target nucleic acid sequence, such as poly-(L-lysine). Further still, intercalating agents, such as ellipticine, and alkylating agents or metal complexes may be attached to sense or antisense oligonucleotides to modify binding specificities of the antisense or sense oligonucleotide for the target nucleotide sequence.

Antisense or sense oligonucleotides may be introduced into a cell containing the target nucleic acid sequence by any gene transfer method, including, for example, CaPO_4 -mediated DNA transfection, electroporation, or by using gene transfer vectors such as Epstein-Barr virus. In a preferred procedure, an antisense or sense oligonucleotide is inserted into a suitable retroviral vector. A cell containing the target nucleic acid sequence is contacted with the recombinant retroviral vector, either *in vivo* or *ex vivo*. Suitable retroviral vectors include, but are not limited to, those derived from the murine retrovirus M-MuLV, N2 (a retrovirus derived from M-MuLV), or the double copy vectors designated DCT5A, DCT5B and DCT5C (see WO 90/13641).

Sense or antisense oligonucleotides also may be introduced into a cell containing the target nucleotide sequence by formation of a conjugate with a ligand binding molecule, as described in WO 91/04753. Suitable ligand binding molecules include, but are not limited to, cell surface receptors, growth factors, other cytokines, or other ligands that bind to cell surface receptors. Preferably, conjugation of the ligand binding molecule does not substantially interfere with the ability of the ligand binding molecule to bind to its corresponding molecule or receptor, or block entry of the sense or antisense oligonucleotide or its conjugated version into the cell.

Alternatively, a sense or an antisense oligonucleotide may be introduced into a cell containing the target nucleic acid sequence by formation of an oligonucleotide-lipid complex, as described in WO 90/10448. The sense or antisense oligonucleotide-lipid complex is preferably dissociated within the cell by an endogenous lipase.

Antisense or sense RNA or DNA molecules are generally at least about 5 bases in length, about 10 bases in length, about 15 bases in length, about 20 bases in length, about 25 bases in length, about 30 bases in length, about 35 bases in length, about 40 bases in length, about 45 bases in length, about 50 bases in length, about 55 bases in length, about 60 bases in length, about 65 bases in length, about 70 bases in length, about 75 bases in length, about 80 bases in length, about 85 bases in length, about 90 bases in length, about 95 bases in length, about 100 bases in length, or more.

The probes may also be employed in PCR techniques to generate a pool of sequences for identification of closely related PRO coding sequences.

Nucleotide sequences encoding a PRO can also be used to construct hybridization probes for mapping the gene which encodes that PRO and for the genetic analysis of individuals with genetic disorders. The nucleotide sequences provided herein may be mapped to a chromosome and specific regions of a chromosome using known techniques, such as *in situ* hybridization, linkage analysis against known chromosomal markers, and hybridization screening with libraries.

When the coding sequences for PRO encode a protein which binds to another protein (example, where the PRO is a receptor), the PRO can be used in assays to identify the other proteins or molecules involved in the binding interaction. By such methods, inhibitors of the receptor/ligand binding interaction can be identified. Proteins involved in such binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction. Also, the receptor PRO can be used to isolate correlative ligand(s). Screening assays can be designed to find lead compounds that mimic the biological activity of a native PRO or a receptor for PRO. Such screening assays will include assays amenable to high-throughput screening of chemical libraries, making them particularly suitable for identifying small molecule drug candidates. Small

molecules contemplated include synthetic organic or inorganic compounds. The assays can be performed in a variety of formats, including protein-protein binding assays, biochemical screening assays, immunoassays and cell based assays, which are well characterized in the art.

Nucleic acids which encode PRO or its modified forms can also be used to generate either transgenic animals or "knock out" animals which, in turn, are useful in the development and screening of therapeutically useful reagents. A transgenic animal (e.g., a mouse or rat) is an animal having cells that contain a transgene, which transgene was introduced into the animal or an ancestor of the animal at a prenatal, e.g., an embryonic stage. A transgene is a DNA which is integrated into the genome of a cell from which a transgenic animal develops. In one embodiment, cDNA encoding PRO can be used to clone genomic DNA encoding PRO in accordance with established techniques and the genomic sequences used to generate transgenic animals that contain cells which express DNA encoding PRO. Methods for generating transgenic animals, particularly animals such as mice or rats, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009. Typically, particular cells would be targeted for PRO transgene incorporation with tissue-specific enhancers. Transgenic animals that include a copy of a transgene encoding PRO introduced into the germ line of the animal at an embryonic stage can be used to examine the effect of increased expression of DNA encoding PRO. Such animals can be used as tester animals for reagents thought to confer protection from, for example, pathological conditions associated with its overexpression. In accordance with this facet of the invention, an animal is treated with the reagent and a reduced incidence of the pathological condition, compared to untreated animals bearing the transgene, would indicate a potential therapeutic intervention for the pathological condition.

Alternatively, non-human homologues of PRO can be used to construct a PRO "knock out" animal which has a defective or altered gene encoding PRO as a result of homologous recombination between the endogenous gene encoding PRO and altered genomic DNA encoding PRO introduced into an embryonic stem cell of the animal. For example, cDNA encoding PRO can be used to clone genomic DNA encoding PRO in accordance with established techniques. A portion of the genomic DNA encoding PRO can be deleted or replaced with another gene, such as a gene encoding a selectable marker which can be used to monitor integration. Typically, several kilobases of unaltered flanking DNA (both at the 5' and 3' ends) are included in the vector [see e.g., Thomas and Capecchi, *Cell*, 51:503 (1987) for a description of homologous recombination vectors]. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced DNA has homologously recombined with the endogenous DNA are selected [see e.g., Li et al., *Cell*, 69:915 (1992)]. The selected cells are then injected into a blastocyst of an animal (e.g., a mouse or rat) to form aggregation chimeras [see e.g., Bradley, in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E. J. Robertson, ed. (IRL, Oxford, 1987), pp. 113-152]. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term to create a "knock out" animal. Progeny harboring the homologously recombined DNA in their germ cells can be identified by standard techniques and used to breed animals in which all cells of the animal contain the homologously recombined DNA. Knockout animals can be characterized for instance, for their ability to defend against certain pathological conditions and for their development of pathological conditions due to absence of

the PRO polypeptide.

Nucleic acid encoding the PRO polypeptides may also be used in gene therapy. In gene therapy applications, genes are introduced into cells in order to achieve *in vivo* synthesis of a therapeutically effective genetic product, for example for replacement of a defective gene. "Gene therapy" includes both conventional gene therapy where a lasting effect is achieved by a single treatment, and the administration of gene therapeutic agents, which involves the one time or repeated administration of a therapeutically effective DNA or mRNA. Antisense RNAs and DNAs can be used as therapeutic agents for blocking the expression of certain genes *in vivo*. It has already been shown that short antisense oligonucleotides can be imported into cells where they act as inhibitors, despite their low intracellular concentrations caused by their restricted uptake by the cell membrane. (Zamecnik *et al.*, Proc. Natl. Acad. Sci. USA 83:4143-4146 [1986]). The oligonucleotides can be modified to enhance their uptake, e.g. by substituting their negatively charged phosphodiester groups by uncharged groups.

There are a variety of techniques available for introducing nucleic acids into viable cells. The techniques vary depending upon whether the nucleic acid is transferred into cultured cells *in vitro*, or *in vivo* in the cells of the intended host. Techniques suitable for the transfer of nucleic acid into mammalian cells *in vitro* include the use of liposomes, electroporation, microinjection, cell fusion, DEAE-dextran, the calcium phosphate precipitation method, etc. The currently preferred *in vivo* gene transfer techniques include transfection with viral (typically retroviral) vectors and viral coat protein-liposome mediated transfection (Dzau *et al.*, Trends in Biotechnology 11, 205-210 [1993]). In some situations it is desirable to provide the nucleic acid source with an agent that targets the target cells, such as an antibody specific for a cell surface membrane protein or the target cell, a ligand for a receptor on the target cell, etc. Where liposomes are employed, proteins which bind to a cell surface membrane protein associated with endocytosis may be used for targeting and/or to facilitate uptake, e.g. capsid proteins or fragments thereof tropic for a particular cell type, antibodies for proteins which undergo internalization in cycling, proteins that target intracellular localization and enhance intracellular half-life. The technique of receptor-mediated endocytosis is described, for example, by Wu *et al.*, J. Biol. Chem. 262, 25 4429-4432 (1987); and Wagner *et al.*, Proc. Natl. Acad. Sci. USA 87, 3410-3414 (1990). For review of gene marking and gene therapy protocols see Anderson *et al.*, Science 256, 808-813 (1992).

The PRO polypeptides described herein may also be employed as molecular weight markers for protein electrophoresis purposes and the isolated nucleic acid sequences may be used for recombinantly expressing those markers.

30 The nucleic acid molecules encoding the PRO polypeptides or fragments thereof described herein are useful for chromosome identification. In this regard, there exists an ongoing need to identify new chromosome markers, since relatively few chromosome marking reagents, based upon actual sequence data are presently available. Each PRO nucleic acid molecule of the present invention can be used as a chromosome marker.

35 The PRO polypeptides and nucleic acid molecules of the present invention may also be used for tissue typing, wherein the PRO polypeptides of the present invention may be differentially expressed in one tissue as compared to another. PRO nucleic acid molecules will find use for generating probes for PCR, Northern analysis, Southern analysis and Western analysis.

The PRO polypeptides described herein may also be employed as therapeutic agents. The PRO polypeptides of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby the PRO product hereof is combined in admixture with a pharmaceutically acceptable carrier vehicle. Therapeutic formulations are prepared for storage by mixing the active ingredient having the desired degree of purity with optional physiologically acceptable carriers, excipients or stabilizers

5 (Remington's Pharmaceutical Sciences 16th edition, Osol, A. Ed. (1980)), in the form of lyophilized formulations or aqueous solutions. Acceptable carriers, excipients or stabilizers are nontoxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate and other organic acids; antioxidants including ascorbic acid; low molecular weight (less than about 10 residues) polypeptides; proteins, such as serum albumin, gelatin or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone, amino

10 acids such as glycine, glutamine, asparagine, arginine or lysine; monosaccharides, disaccharides and other carbohydrates including glucose, mannose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; salt-forming counterions such as sodium; and/or nonionic surfactants such as TWEEN™, PLURONICS™ or PEG.

20 The formulations to be used for *in vivo* administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes, prior to or following lyophilization and reconstitution.

25 Therapeutic compositions herein generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

The route of administration is in accord with known methods, e.g. injection or infusion by intravenous, intraperitoneal, intracerebral, intramuscular, intraocular, intraarterial or intralesional routes, topical administration, or by sustained release systems.

30 Dosages and desired drug concentrations of pharmaceutical compositions of the present invention may vary depending on the particular use envisioned. The determination of the appropriate dosage or route of administration is well within the skill of an ordinary physician. Animal experiments provide reliable guidance for the determination of effective doses for human therapy. Interspecies scaling of effective doses can be performed following the principles laid down by Mordenti, J. and Chappell, W. "The use of interspecies scaling in toxicokinetics" In Toxicokinetics and New Drug Development, Yacobi et al., Eds., Pergamon Press, New York 1989, pp. 42-96.

35 When *in vivo* administration of a PRO polypeptide or agonist or antagonist thereof is employed, normal dosage amounts may vary from about 10 ng/kg to up to 100 mg/kg of mammal body weight or more per day, preferably about 1 μ g/kg/day to 10 mg/kg/day, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature; see, for example, U.S. Pat. Nos. 4,657,760; 5,206,344; or 5,225,212. It is anticipated that different formulations will be effective for different treatment compounds and different disorders, that administration targeting one organ or tissue, for example, may necessitate delivery in a manner different from that to another organ or tissue.

35 Where sustained-release administration of a PRO polypeptide is desired in a formulation with release characteristics suitable for the treatment of any disease or disorder requiring administration of the PRO polypeptide, microencapsulation of the PRO polypeptide is contemplated. Microencapsulation of recombinant

proteins for sustained release has been successfully performed with human growth hormone (rhGH), interferon- (rhIFN-), interleukin-2, and MN rgp120. Johnson et al., Nat. Med., 2:795-799 (1996); Yasuda, Biomed. Ther., 27:1221-1223 (1993); Hora et al., Bio/Technology, 8:755-758 (1990); Cleland, "Design and Production of Single Immunization Vaccines Using Polylactide Polyglycolide Microsphere Systems," in Vaccine Design: The Subunit and Adjuvant Approach, Powell and Newman, eds, (Plenum Press: New York, 1995), pp. 439-462;

5 WO 97/03692, WO 96/40072, WO 96/07399; and U.S. Pat. No. 5,654,010.

The sustained-release formulations of these proteins were developed using poly-lactic-coglycolic acid (PLGA) polymer due to its biocompatibility and wide range of biodegradable properties. The degradation products of PLGA, lactic and glycolic acids, can be cleared quickly within the human body. Moreover, the degradability of this polymer can be adjusted from months to years depending on its molecular weight and composition. Lewis, "Controlled release of bioactive agents from lactide/glycolide polymer," in: M. Chasin and R. Langer (Eds.), Biodegradable Polymers as Drug Delivery Systems (Marcel Dekker: New York, 1990), pp. 1-41.

This invention encompasses methods of screening compounds to identify those that mimic the PRO polypeptide (agonists) or prevent the effect of the PRO polypeptide (antagonists). Screening assays for antagonist drug candidates are designed to identify compounds that bind or complex with the PRO polypeptides encoded by the genes identified herein, or otherwise interfere with the interaction of the encoded polypeptides with other cellular proteins. Such screening assays will include assays amenable to high-throughput screening of chemical libraries, making them particularly suitable for identifying small molecule drug candidates.

The assays can be performed in a variety of formats, including protein-protein binding assays, biochemical screening assays, immunoassays, and cell-based assays, which are well characterized in the art.

All assays for antagonists are common in that they call for contacting the drug candidate with a PRO polypeptide encoded by a nucleic acid identified herein under conditions and for a time sufficient to allow these two components to interact.

In binding assays, the interaction is binding and the complex formed can be isolated or detected in the reaction mixture. In a particular embodiment, the PRO polypeptide encoded by the gene identified herein or the drug candidate is immobilized on a solid phase, e.g., on a microtiter plate, by covalent or non-covalent attachments. Non-covalent attachment generally is accomplished by coating the solid surface with a solution of the PRO polypeptide and drying. Alternatively, an immobilized antibody, e.g., a monoclonal antibody, specific for the PRO polypeptide to be immobilized can be used to anchor it to a solid surface. The assay is performed by adding the non-immobilized component, which may be labeled by a detectable label, to the immobilized component, e.g., the coated surface containing the anchored component. When the reaction is complete, the non-reacted components are removed, e.g., by washing, and complexes anchored on the solid surface are detected. When the originally non-immobilized component carries a detectable label, the detection of label immobilized on the surface indicates that complexing occurred. Where the originally non-immobilized component does not carry a label, complexing can be detected, for example, by using a labeled antibody specifically binding the immobilized complex.

If the candidate compound interacts with but does not bind to a particular PRO polypeptide encoded by a gene identified herein, its interaction with that polypeptide can be assayed by methods well known for detecting protein-protein interactions. Such assays include traditional approaches, such as, e.g., cross-linking, co-immunoprecipitation, and co-purification through gradients or chromatographic columns. In addition, protein-protein interactions can be monitored by using a yeast-based genetic system described by Fields and co-workers (Fields and Song, Nature (London), 340:245-246 (1989); Chien et al., Proc. Natl. Acad. Sci. USA, 88:9578-9582 (1991)) as disclosed by Chevray and Nathans, Proc. Natl. Acad. Sci. USA, 89: 5789-5793 (1991). Many transcriptional activators, such as yeast GAL4, consist of two physically discrete modular domains, one acting as the DNA-binding domain, the other one functioning as the transcription-activation domain. The yeast expression system described in the foregoing publications (generally referred to as the "two-hybrid system") takes advantage of this property, and employs two hybrid proteins, one in which the target protein is fused to the DNA-binding domain of GAL4, and another, in which candidate activating proteins are fused to the activation domain. The expression of a GAL1-*lacZ* reporter gene under control of a GAL4-activated promoter depends on reconstitution of GAL4 activity via protein-protein interaction. Colonies containing interacting polypeptides are detected with a chromogenic substrate for β -galactosidase. A complete kit (MATCHMAKERTM) for identifying protein-protein interactions between two specific proteins using the two-hybrid technique is commercially available from Clontech. This system can also be extended to map protein domains involved in specific protein interactions as well as to pinpoint amino acid residues that are crucial for these interactions.

Compounds that interfere with the interaction of a gene encoding a PRO polypeptide identified herein and other intra- or extracellular components can be tested as follows: usually a reaction mixture is prepared containing the product of the gene and the intra- or extracellular component under conditions and for a time allowing for the interaction and binding of the two products. To test the ability of a candidate compound to inhibit binding, the reaction is run in the absence and in the presence of the test compound. In addition, a placebo may be added to a third reaction mixture, to serve as positive control. The binding (complex formation) between the test compound and the intra- or extracellular component present in the mixture is monitored as described hereinabove. The formation of a complex in the control reaction(s) but not in the reaction mixture containing the test compound indicates that the test compound interferes with the interaction of the test compound and its reaction partner.

To assay for antagonists, the PRO polypeptide may be added to a cell along with the compound to be screened for a particular activity and the ability of the compound to inhibit the activity of interest in the presence of the PRO polypeptide indicates that the compound is an antagonist to the PRO polypeptide. Alternatively, antagonists may be detected by combining the PRO polypeptide and a potential antagonist with membrane-bound PRO polypeptide receptors or recombinant receptors under appropriate conditions for a competitive inhibition assay. The PRO polypeptide can be labeled, such as by radioactivity, such that the number of PRO polypeptide molecules bound to the receptor can be used to determine the effectiveness of the potential antagonist. The gene encoding the receptor can be identified by numerous methods known to those of skill in the art, for example, ligand panning and FACS sorting. Coligan et al., Current Protocols in Immun., 1(2): Chapter 5 (1991).

Preferably, expression cloning is employed wherein polyadenylated RNA is prepared from a cell responsive to the PRO polypeptide and a cDNA library created from this RNA is divided into pools and used to transfect COS cells or other cells that are not responsive to the PRO polypeptide. Transfected cells that are grown on glass slides are exposed to labeled PRO polypeptide. The PRO polypeptide can be labeled by a variety of means including iodination or inclusion of a recognition site for a site-specific protein kinase. Following fixation and 5 incubation, the slides are subjected to autoradiographic analysis. Positive pools are identified and sub-pools are prepared and re-transfected using an interactive sub-pooling and re-screening process, eventually yielding a single clone that encodes the putative receptor.

As an alternative approach for receptor identification, labeled PRO polypeptide can be photoaffinity-linked with cell membrane or extract preparations that express the receptor molecule. Cross-linked material is 10 resolved by PAGE and exposed to X-ray film. The labeled complex containing the receptor can be excised, resolved into peptide fragments, and subjected to protein micro-sequencing. The amino acid sequence obtained from micro- sequencing would be used to design a set of degenerate oligonucleotide probes to screen a cDNA library to identify the gene encoding the putative receptor.

15 In another assay for antagonists, mammalian cells or a membrane preparation expressing the receptor would be incubated with labeled PRO polypeptide in the presence of the candidate compound. The ability of the compound to enhance or block this interaction could then be measured.

20 More specific examples of potential antagonists include an oligonucleotide that binds to the fusions of immunoglobulin with PRO polypeptide, and, in particular, antibodies including, without limitation, poly- and monoclonal antibodies and antibody fragments, single-chain antibodies, anti-idiotypic antibodies, and chimeric or humanized versions of such antibodies or fragments, as well as human antibodies and antibody fragments. Alternatively, a potential antagonist may be a closely related protein, for example, a mutated form of the PRO polypeptide that recognizes the receptor but imparts no effect, thereby competitively inhibiting the action of the PRO polypeptide.

25 Another potential PRO polypeptide antagonist is an antisense RNA or DNA construct prepared using antisense technology, where, e.g., an antisense RNA or DNA molecule acts to block directly the translation of mRNA by hybridizing to targeted mRNA and preventing protein translation. Antisense technology can be used to control gene expression through triple-helix formation or antisense DNA or RNA, both of which methods are based on binding of a polynucleotide to DNA or RNA. For example, the 5' coding portion of the polynucleotide sequence, which encodes the mature PRO polypeptides herein, is used to design an antisense RNA 30 oligonucleotide of from about 10 to 40 base pairs in length. A DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res., 6:3073 (1979); Cooney et al., Science, 241: 456 (1988); Dervan et al., Science, 251:1360 (1991)), thereby preventing transcription and the production of the PRO polypeptide. The antisense RNA oligonucleotide hybridizes to the mRNA *in vivo* and blocks translation of the mRNA molecule into the PRO polypeptide 35 (antisense - Okano, Neurochem., 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression (CRC Press: Boca Raton, FL, 1988). The oligonucleotides described above can also be delivered to cells such that the antisense RNA or DNA may be expressed *in vivo* to inhibit production of the PRO

polypeptide. When antisense DNA is used, oligodeoxyribonucleotides derived from the translation-initiation site, e.g., between about -10 and +10 positions of the target gene nucleotide sequence, are preferred.

Potential antagonists include small molecules that bind to the active site, the receptor binding site, or growth factor or other relevant binding site of the PRO polypeptide, thereby blocking the normal biological activity of the PRO polypeptide. Examples of small molecules include, but are not limited to, small peptides 5 or peptide-like molecules, preferably soluble peptides, and synthetic non-peptidyl organic or inorganic compounds.

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. Ribozymes act by sequence-specific hybridization to the complementary target RNA, followed by endonucleolytic cleavage. Specific ribozyme cleavage sites within a potential RNA target can be identified by known techniques. 10 For further details see, e.g., Rossi, Current Biology, 4:469-471 (1994), and PCT publication No. WO 97/33551 (published September 18, 1997).

Nucleic acid molecules in triple-helix formation used to inhibit transcription should be single-stranded and composed of deoxynucleotides. The base composition of these oligonucleotides is designed such that it promotes triple-helix formation via Hoogsteen base-pairing rules, which generally require sizeable stretches of purines or pyrimidines on one strand of a duplex. For further details see, e.g., PCT publication No. WO 97/33551, *supra*.

These small molecules can be identified by any one or more of the screening assays discussed hereinabove and/or by any other screening techniques well known for those skilled in the art.

PRO241 polypeptides of the present invention which possess biological activity related to that of the endogenous biglycan protein may be employed both *in vivo* for therapeutic purposes and *in vitro*. Those of ordinary skill in the art will well know how to employ the PRO241 polypeptides of the present invention for such purposes.

Chordin is a candidate gene for a dysmorphia syndrome known as Cornelia de Lange Syndrome (CDL) which is characterized by distinctive facial features (low anterior hairline, synophrys, anteneted nares, maxillary 25 prognathism, long philtrum, 'carp' mouth), prenatal and postnatal growth retardation, mental retardation and, often but not always, upper limb abnormalities. There are also rare cases where CDL is present in association with thrombocytopenia. The gene for CDL has been mapped by linkage to 3q26.3 (OMIM #122470). Xchd involvement in early *Xenopus* patterning and nervous system development makes CHD an intriguing candidate gene. CHD maps to the appropriate region on chromosome 3. It is very close to THPO, and deletions 30 encompassing both THPO and CHD could result in rare cases of thrombocytopenia and developmental abnormalities. *In situ* analysis of CD revealed that almost all adult tissues are negative for CHD expression, the only positive signal was observed in the cleavage line of the developing synovial joint forming between the femoral head and acetabulum (hip joint) implicating CHD in the development and presumably growth of long bones. Such a function, if disrupted, could result in growth retardation.

35 The human CHD amino acid sequence predicted from the cDNA is 50% identical (and 66% conserved) to Xchd. All 40 cysteines in the 4 cysteine-rich domains are conserved. These cysteine rich domains are similar to those observed in thrombospondin, procollagen and von Willebrand factor. Bornstein, P. FASEB J 6: 3290-

3299 (1992); Hunt, L. & Barker, W. *Biochem. Biophys. Res. Commun.* 144: 876-882 (1987).

The human CHD locus (genomic PRO243) comprises 23 exons in 9.6 kb of genomic DNA. The initiating methionine is in exon 1 and the stop codon in exon 23. A CpG island is located at the 5' end of the gene, beginning approximately 100 bp 5' of exon 1 and extends through the first exon and ends within the first intron. The THPO and CHD loci are organized in a head-to-head fashion with approximately 2.2 kb separating 5 their transcription start sites. At the protein level, PRO243 is 51% identical to Xenopus chordin (Xchd). All forty cysteines in the one amino terminal and three carboxy terminal cysteine-rich clusters are conserved.

PRO243 is a 954 amino acid polypeptide having a signal sequence at residues 1 to about 23. There are 4 cysteine clusters: (1) residues about 51 to about 125; (2) residues about 705 to about 761; (3) residues about 784 to about 849; and (4) residues about 897 to about 931. There are potential leucine zippers at residues about 10 315 to about 396, and N-glycosylation sites at residues 217, 351, 365 and 434.

PRO299 polypeptides and portions thereof which have homology to the notch protein may be useful for *in vivo* therapeutic purposes, as well as for various other applications. The identification of novel notch proteins and related molecules may be relevant to a number of human disorders such as those effecting development. Thus, the identification of new notch proteins and notch-like molecules is of special importance in that such 15 proteins may serve as potential therapeutics for a variety of different human disorders. Such polypeptides may also play important roles in biotechnological and medical research as well as various industrial applications. As a result, there is particular scientific and medical interest in new molecules, such as PRO299.

PRO323 polypeptides of the present invention which possess biological activity related to that of one or more endogenous dipeptidase proteins may be employed both *in vivo* for therapeutic purposes and *in vitro*. Those of ordinary skill in the art will well know how to employ the PRO323 polypeptides of the present 20 invention for such purposes.

PRO327 polypeptides of the present invention which possess biological activity related to that of the endogenous prolactin receptor protein may be employed both *in vivo* for therapeutic purposes and *in vitro*. Those of ordinary skill in the art will well know how to employ the PRO327 polypeptides of the present 25 invention for such purposes. PRO327 polypeptides which possess the ability to bind to prolactin may function both *in vitro* and *in vivo* as prolactin antagonists.

PRO233 polypeptides and portions thereof which have homology to reductase may also be useful for *in vivo* therapeutic purposes, as well as for various other applications. The identification of novel reductase proteins and related molecules may be relevant to a number of human disorders such as inflammatory disease, 30 organ failure, atherosclerosis, cardiac injury, infertility, birth defects, premature aging, AIDS, cancer, diabetic complications and mutations in general. Given that oxygen free radicals and antioxidants appear to play important roles in a number of disease processes, the identification of new reductase proteins and reductase-like molecules is of special importance in that such proteins may serve as potential therapeutics for a variety of different human disorders. Such polypeptides may also play important roles in biotechnological and medical 35 research, as well as various industrial applications. As a result, there is particular scientific and medical interest in new molecules, such as PRO233.

PRO344 polypeptides and portions thereof which have homology to complement proteins may also be useful for *in vivo* therapeutic purposes, as well as for various other applications. The identification of novel complement proteins and related molecules may be relevant to a number of human disorders such as effecting the inflammatory response of cells of the immune system. Thus, the identification of new complement proteins and complement-like molecules is of special importance in that such proteins may serve as potential therapeutics for a variety of different human disorders. Such polypeptides may also play important roles in biotechnological and medical research as well as various industrial applications. As a result, there is particular scientific and medical interest in new molecules, such as PRO344.

PRO347 polypeptides of the present invention which possess biological activity related to that of cysteine-rich secretory proteins may be employed both *in vivo* for therapeutic purposes and *in vitro*. Those of ordinary skill in the art will well know how to employ the PRO347 polypeptides of the present invention for such purposes.

PRO354 polypeptides of the present invention which possess biological activity related to that of the heavy chain of the inter-alpha-trypsin inhibitor protein may be employed both *in vivo* for therapeutic purposes and *in vitro*. Those of ordinary skill in the art will well know how to employ the PRO354 polypeptides of the present invention for such purposes.

PRO355 polypeptides and portions thereof which have homology to CRTAM may also be useful for *in vivo* therapeutic purposes, as well as for various other applications. The identification of novel molecules associated with T cells may be relevant to a number of human disorders such as conditions involving the immune system in general. Given that the CRTAM protein binds antibodies which play important roles in a number of disease processes, the identification of new CRTAM proteins and CRTAM-like molecules is of special importance in that such proteins may serve as potential therapeutics for a variety of different human disorders. Such polypeptides may also play important roles in biotechnological and medical research, as well as various industrial applications. As a result, there is particular scientific and medical interest in new molecules, such as PRO355.

PRO357 can be used in competitive binding assays with ALS to determine its activity with respect to ALS. Moreover, PRO357 can be used in assays to determine if it prolongs polypeptides which it may complex with to have longer half-lives in vivo. PRO357 can be used similarly in assays with carboxypeptidase, to which it also has homology. The results can be applied accordingly.

PRO715 polypeptides of the present invention which possess biological activity related to that of the tumor necrosis factor family of proteins may be employed both *in vivo* for therapeutic purposes and *in vitro*. Those of ordinary skill in the art will well know how to employ the PRO715 polypeptides of the present invention for such purposes. PRO715 polypeptides will be expected to bind to their specific receptors, thereby activating such receptors. Variants of the PRO715 polypeptides of the present invention may function as agonists or antagonists of their specific receptor activity.

PRO353 polypeptides and portions thereof which have homology to the complement protein may also be useful for *in vivo* therapeutic purposes, as well as for various other applications. The identification of novel complement proteins and related molecules may be relevant to a number of human disorders such as effecting

the inflammatory response of cells of the immune system. Thus, the identification of new complement proteins complement-like molecules is of special importance in that such proteins may serve as potential therapeutics for a variety of different human disorders. Such polypeptides may also play important roles in biotechnological and medical research as well as various industrial applications. As a result, there is particular scientific and medical interest in new molecules, such as PRO353.

5 PRO361 polypeptides and portions thereof which have homology to mucin and/or chitinase proteins may also be useful for *in vivo* therapeutic purposes, as well as for various other applications. The identification of novel mucin and/or chitinase proteins and related molecules may be relevant to a number of human disorders such as cancer or those involving cell surface molecules or receptors. Thus, the identification of new mucin and/or chitinase proteins is of special importance in that such proteins may serve as potential therapeutics for
10 a variety of different human disorders. Such polypeptides may also play important roles in biotechnological and medical research as well as various industrial applications. As a result, there is particular scientific and medical interest in new molecules, such as PRO361.

15 PRO365 polypeptides and portions thereof which have homology to the human 2-19 protein may also be useful for *in vivo* therapeutic purposes, as well as for various other applications. The identification of novel human 2-19 proteins and related molecules may be relevant to a number of human disorders such as modulating the binding or activity of cells of the immune system. Thus, the identification of new human 2-19 proteins and human 2-19 protein-like molecules is of special importance in that such proteins may serve as potential therapeutics for a variety of different human disorders. Such polypeptides may also play important roles in biotechnological and medical research as well as various industrial applications. As a result, there is particular
20 scientific and medical interest in new molecules, such as PRO365.

F. Anti-PRO Antibodies

25 The present invention further provides anti-PRO antibodies. Exemplary antibodies include polyclonal, monoclonal, humanized, bispecific, and heteroconjugate antibodies.

1. Polyclonal Antibodies

30 The anti-PRO antibodies may comprise polyclonal antibodies. Methods of preparing polyclonal antibodies are known to the skilled artisan. Polyclonal antibodies can be raised in a mammal, for example, by one or more injections of an immunizing agent and, if desired, an adjuvant. Typically, the immunizing agent and/or adjuvant will be injected in the mammal by multiple subcutaneous or intraperitoneal injections. The immunizing agent may include the PRO polypeptide or a fusion protein thereof. It may be useful to conjugate the immunizing agent to a protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. Examples of adjuvants which may be employed include Freund's complete adjuvant and MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).
35 The immunization protocol may be selected by one skilled in the art without undue experimentation.

2. Monoclonal Antibodies

The anti-PRO antibodies may, alternatively, be monoclonal antibodies. Monoclonal antibodies may be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes may be immunized *in vitro*.

The immunizing agent will typically include the PRO polypeptide or a fusion protein thereof. Generally, either peripheral blood lymphocytes ("PBLs") are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell [Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103]. Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells may be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies [Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63].

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against PRO. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an *in vitro* binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980).

After the desired hybridoma cells are identified, the clones may be subcloned by limiting dilution procedures and grown by standard methods [Goding, supra]. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells may be grown *in vivo* as ascites in a mammal.

The monoclonal antibodies secreted by the subclones may be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies may also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA may be placed into expression 5 vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also may be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences [U.S. Patent No. 4,816,567; Morrison et al., *supra*] or by covalently joining to the immunoglobulin coding 10 sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

15 The antibodies may be monovalent antibodies. Methods for preparing monovalent antibodies are well known in the art. For example, one method involves recombinant expression of immunoglobulin light chain and modified heavy chain. The heavy chain is truncated generally at any point in the Fc region so as to prevent heavy chain crosslinking. Alternatively, the relevant cysteine residues are substituted with another amino acid residue or are deleted so as to prevent crosslinking.

20 *In vitro* methods are also suitable for preparing monovalent antibodies. Digestion of antibodies to produce fragments thereof, particularly, Fab fragments, can be accomplished using routine techniques known in the art.

3. Human and Humanized Antibodies

25 The anti-PRO antibodies of the invention may further comprise humanized antibodies or human antibodies. Humanized forms of non-human (e.g., murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')₂ or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. Humanized antibodies include human immunoglobulins (recipient antibody) in which residues from a 30 complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies may also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR 35 regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin [Jones

et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-329 (1988); and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)].

Methods for humanizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. These non-human amino acid residues are often referred to as "import" residues, which are typically taken from an "import" variable domain. Humanization can be essentially performed following the method of Winter and co-workers [Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)], by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. Accordingly, such "humanized" antibodies are chimeric antibodies (U.S. Patent No. 4,816,567), wherein substantially less than an intact human variable domain has been substituted by the corresponding sequence from a non-human species. In practice, humanized antibodies are typically human antibodies in which some CDR residues and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies.

Human antibodies can also be produced using various techniques known in the art, including phage display libraries [Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)]. The techniques of Cole et al. and Boerner et al. are also available for the preparation of human monoclonal antibodies (Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, p. 77 (1985) and Boerner et al., J. Immunol., 147(1):86-95 (1991)]. Similarly, human antibodies can be made by introducing of human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in the following scientific publications: Marks *et al.*, Bio/Technology 10, 779-783 (1992); Lonberg *et al.*, Nature 368 856-859 (1994); Morrison, Nature 368, 812-13 (1994); Fishwild *et al.*, Nature Biotechnology 14, 845-51 (1996); Neuberger, Nature Biotechnology 14, 826 (1996); Lonberg and Huszar, Intern. Rev. Immunol. 13 65-93 (1995).

4. Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for the PRO, the other one is for any other antigen, and preferably for a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities [Milstein and Cuello, Nature, 305:537-539 (1983)]. Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps.

Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Fab' fragments may be directly recovered from *E. coli* and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from *E. coli* and subjected to directed chemical coupling *in vitro* to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various technique for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers

were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger *et al.*, Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the 5 two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber *et al.*, J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. 10 Tutt *et al.*, J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies may bind to two different epitopes on a given PRO polypeptide herein. Alternatively, an anti-PRO polypeptide arm may be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (Fc γ R), such as Fc γ RI (CD64), Fc γ RII (CD32) and Fc γ RIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular PRO polypeptide. Bispecific antibodies may also be used to localize cytotoxic agents to cells which express a particular PRO polypeptide. These antibodies possess a PRO-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the PRO polypeptide and further binds tissue factor (TF).

5. Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells [U.S. Patent No. 4,676,980], and for treatment of HIV infection 25 [WO 91/00360; WO 92/200373; EP 03089]. It is contemplated that the antibodies may be prepared *in vitro* using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins may be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptopbutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

30 6. Effector Function Engineering

It may be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, *e.g.*, the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) may be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The 35 homodimeric antibody thus generated may have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron *et al.*, J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies

with enhanced anti-tumor activity may also be prepared using heterobifunctional cross-linkers as described in Wolff *et al.* Cancer Research, **53**: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and may thereby have enhanced complement lysis and ADCC capabilities. See Stevenson *et al.*, Anti-Cancer Drug Design, 3: 219-230 (1989).

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (*e.g.*, an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (*i.e.*, a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above.

10 Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, *Aleurites fordii* proteins, dianthin proteins, *Phytolaca americana* proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the trichothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi , ^{131}I , ^{131}In , ^{90}Y , and ^{186}Re .

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridylthiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta *et al.*, Science, **238**: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyl diethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

25 In another embodiment, the antibody may be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is conjugated to a cytotoxic agent (e.g., a radionuclide).

30 8. Immunoliposomes

The antibodies disclosed herein may also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein *et al.*, Proc. Natl. Acad. Sci. USA, 82: 3688 (1985); Hwang *et al.*, Proc. Natl. Acad. Sci. USA, 77: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-).

PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present invention can be conjugated to the liposomes as described in Martin *et al.*, J. Biol. Chem., 257: 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as Doxorubicin) is optionally contained within the liposome. See Gabizon *et al.*, J. National Cancer Inst., 81(19): 1484 (1989).

5

9. Pharmaceutical Compositions of Antibodies

Antibodies specifically binding a PRO polypeptide identified herein, as well as other molecules identified by the screening assays disclosed hereinbefore, can be administered for the treatment of various disorders in the form of pharmaceutical compositions.

10 If the PRO polypeptide is intracellular and whole antibodies are used as inhibitors, internalizing antibodies are preferred. However, lipofections or liposomes can also be used to deliver the antibody, or an antibody fragment, into cells. Where antibody fragments are used, the smallest inhibitory fragment that specifically binds to the binding domain of the target protein is preferred. For example, based upon the variable-region sequences of an antibody, peptide molecules can be designed that retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant DNA technology. See, *e.g.*, Marasco *et al.*, Proc. Natl. Acad. Sci. USA, 90: 7889-7893 (1993). The formulation herein may also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition may comprise an agent that enhances its function, such as, for example, a cytotoxic agent, cytokine, chemotherapeutic agent, or growth-inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended.

15 The active ingredients may also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, 20 liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions. Such techniques are disclosed in Remington's Pharmaceutical Sciences, supra.

25 The formulations to be used for *in vivo* administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

30 Sustained-release preparations may be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, *e.g.*, films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and γ ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT TM (injectable 35 microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods. When encapsulated

antibodies remain in the body for a long time, they may denature or aggregate as a result of exposure to moisture at 37°C, resulting in a loss of biological activity and possible changes in immunogenicity. Rational strategies can be devised for stabilization depending on the mechanism involved. For example, if the aggregation mechanism is discovered to be intermolecular S-S bond formation through thio-disulfide interchange, stabilization may be achieved by modifying sulphydryl residues, lyophilizing from acidic solutions, controlling moisture content, using appropriate additives, and developing specific polymer matrix compositions.

5 G. Uses for anti-PRO Antibodies

The anti-PRO antibodies of the invention have various utilities. For example, anti-PRO antibodies may be used in diagnostic assays for PRO, *e.g.*, detecting its expression in specific cells, tissues, or serum. Various 10 diagnostic assay techniques known in the art may be used, such as competitive binding assays, direct or indirect sandwich assays and immunoprecipitation assays conducted in either heterogeneous or homogeneous phases [Zola, Monoclonal Antibodies: A Manual of Techniques, CRC Press, Inc. (1987) pp. 147-158]. The antibodies used in the diagnostic assays can be labeled with a detectable moiety. The detectable moiety should be capable 15 of producing, either directly or indirectly, a detectable signal. For example, the detectable moiety may be a radioisotope, such as ^3H , ^{14}C , ^{32}P , ^{35}S , or ^{125}I , a fluorescent or chemiluminescent compound, such as fluorescein isothiocyanate, rhodamine, or luciferin, or an enzyme, such as alkaline phosphatase, beta-galactosidase or horseradish peroxidase. Any method known in the art for conjugating the antibody to the detectable moiety may 20 be employed, including those methods described by Hunter et al., Nature, 144:945 (1962); David et al., Biochemistry, 13:1014 (1974); Pain et al., J. Immunol. Meth., 40:219 (1981); and Nygren, J. Histochem. and Cytochem., 30:407 (1982).

Anti-PRO antibodies also are useful for the affinity purification of PRO from recombinant cell culture or natural sources. In this process, the antibodies against PRO are immobilized on a suitable support, such a Sephadex resin or filter paper, using methods well known in the art. The immobilized antibody then is contacted with a sample containing the PRO to be purified, and thereafter the support is washed with a suitable solvent that 25 will remove substantially all the material in the sample except the PRO, which is bound to the immobilized antibody. Finally, the support is washed with another suitable solvent that will release the PRO from the antibody.

The following examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

30 All patent and literature references cited in the present specification are hereby incorporated by reference in their entirety.

EXAMPLES

Commercially available reagents referred to in the examples were used according to manufacturer's 35 instructions unless otherwise indicated. The source of those cells identified in the following examples, and throughout the specification, by ATCC accession numbers is the American Type Culture Collection, Manassas, VA.

EXAMPLE 1: Extracellular Domain Homology Screening to Identify Novel Polypeptides and cDNA Encoding

Therefor

The extracellular domain (ECD) sequences (including the secretion signal sequence, if any) from about 950 known secreted proteins from the Swiss-Prot public database were used to search EST databases. The EST databases included public databases (e.g., Dayhoff, GenBank), and proprietary databases (e.g. LIFESEQ™,

5 Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST-2 (Altschul et al., *Methods in Enzymology* 266:460-480 (1996)) as a comparison of the ECD protein sequences to a 6 frame translation of the EST sequences. Those comparisons with a BLAST score of 70 (or in some cases 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, WA).

10 Using this extracellular domain homology screen, consensus DNA sequences were assembled relative to the other identified EST sequences using phrap. In addition, the consensus DNA sequences obtained were often (but not always) extended using repeated cycles of BLAST or BLAST-2 and phrap to extend the consensus sequence as far as possible using the sources of EST sequences discussed above.

15 Based upon the consensus sequences obtained as described above, oligonucleotides were then synthesized and used to identify by PCR a cDNA library that contained the sequence of interest and for use as probes to isolate a clone of the full-length coding sequence for a PRO polypeptide. Forward and reverse PCR primers generally range from 20 to 30 nucleotides and are often designed to give a PCR product of about 100-1000 bp in length. The probe sequences are typically 40-55 bp in length. In some cases, additional oligonucleotides are synthesized when the consensus sequence is greater than about 1-1.5kbp. In order to screen 20 several libraries for a full-length clone, DNA from the libraries was screened by PCR amplification, as per Ausubel et al., *Current Protocols in Molecular Biology*, with the PCR primer pair. A positive library was then used to isolate clones encoding the gene of interest using the probe oligonucleotide and one of the primer pairs.

25 The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRK5B or pRKD; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., *Science*, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

30 EXAMPLE 2: Isolation of cDNA clones by Amylase Screening

1. Preparation of oligo dT primed cDNA library

mRNA was isolated from a human tissue of interest using reagents and protocols from Invitrogen, San Diego, CA (Fast Track 2). This RNA was used to generate an oligo dT primed cDNA library in the vector pRK5D using reagents and protocols from Life Technologies, Gaithersburg, MD (Super Script Plasmid System).

35 In this procedure, the double stranded cDNA was sized to greater than 1000 bp and the SalI/NotI linkered cDNA was cloned into XhoI/NotI cleaved vector. pRK5D is a cloning vector that has an sp6 transcription initiation site followed by an SfiI restriction enzyme site preceding the XhoI/NotI cDNA cloning sites.

2. Preparation of random primed cDNA library

A secondary cDNA library was generated in order to preferentially represent the 5' ends of the primary cDNA clones. Sp6 RNA was generated from the primary library (described above), and this RNA was used to generate a random primed cDNA library in the vector pSST-AMY.0 using reagents and protocols from Life Technologies (Super Script Plasmid System, referenced above). In this procedure the double stranded cDNA was sized to 500-1000 bp, linkerered with blunt to NotI adaptors, cleaved with SfiI, and cloned into SfiI/NotI cleaved vector. pSST-AMY.0 is a cloning vector that has a yeast alcohol dehydrogenase promoter preceding the cDNA cloning sites and the mouse amylase sequence (the mature sequence without the secretion signal) followed by the yeast alcohol dehydrogenase terminator, after the cloning sites. Thus, cDNAs cloned into this vector that are fused in frame with amylase sequence will lead to the secretion of amylase from appropriately transfected yeast colonies.

3. Transformation and Detection

DNA from the library described in paragraph 2 above was chilled on ice to which was added electrocompetent DH10B bacteria (Life Technologies, 20 ml). The bacteria and vector mixture was then electroporated as recommended by the manufacturer. Subsequently, SOC media (Life Technologies, 1 ml) was added and the mixture was incubated at 37°C for 30 minutes. The transformants were then plated onto 20 standard 150 mm LB plates containing ampicillin and incubated for 16 hours (37°C). Positive colonies were scraped off the plates and the DNA was isolated from the bacterial pellet using standard protocols, e.g. CsCl-gradient. The purified DNA was then carried on to the yeast protocols below.

The yeast methods were divided into three categories: (1) Transformation of yeast with the plasmid/cDNA combined vector; (2) Detection and isolation of yeast clones secreting amylase; and (3) PCR amplification of the insert directly from the yeast colony and purification of the DNA for sequencing and further analysis.

The yeast strain used was HD56-5A (ATCC-90785). This strain has the following genotype: MAT alpha, ura3-52, leu2-3, leu2-112, his3-11, his3-15, MAL⁺, SUC⁺, GAL⁺. Preferably, yeast mutants can be employed that have deficient post-translational pathways. Such mutants may have translocation deficient alleles in *sec71*, *sec72*, *sec62*, with truncated *sec71* being most preferred. Alternatively, antagonists (including antisense nucleotides and/or ligands) which interfere with the normal operation of these genes, other proteins implicated in this post translation pathway (e.g., SEC61p, SEC72p, SEC62p, SEC63p, TDJ1p or SSA1p-4p) or the complex formation of these proteins may also be preferably employed in combination with the amylase-expressing yeast.

Transformation was performed based on the protocol outlined by Gietz et al., Nucl. Acid. Res., 20:1425 (1992). Transformed cells were then inoculated from agar into YEPD complex media broth (100 ml) and grown overnight at 30°C. The YEPD broth was prepared as described in Kaiser et al., Methods in Yeast Genetics, Cold Spring Harbor Press, Cold Spring Harbor, NY, p. 207 (1994). The overnight culture was then diluted to about 2 x 10⁶ cells/ml (approx. OD₆₀₀=0.1) into fresh YEPD broth (500 ml) and regrown to 1 x 10⁷ cells/ml (approx. OD₆₀₀=0.4-0.5).

The cells were then harvested and prepared for transformation by transfer into GS3 rotor bottles in a Sorval GS3 rotor at 5,000 rpm for 5 minutes, the supernatant discarded, and then resuspended into sterile water, and centrifuged again in 50 ml falcon tubes at 3,500 rpm in a Beckman GS-6KR centrifuge. The supernatant was discarded and the cells were subsequently washed with LiAc/TE (10 ml, 10 mM Tris-HCl, 1 mM EDTA pH 7.5, 100 mM Li₂OOCCH₃), and resuspended into LiAc/TE (2.5 ml).

5 Transformation took place by mixing the prepared cells (100 μ l) with freshly denatured single stranded salmon testes DNA (Lofstrand Labs, Gaithersburg, MD) and transforming DNA (1 μ g, vol. < 10 μ l) in microfuge tubes. The mixture was mixed briefly by vortexing, then 40% PEG/TE (600 μ l, 40% polyethylene glycol-4000, 10 mM Tris-HCl, 1 mM EDTA, 100 mM Li₂OOCCH₃, pH 7.5) was added. This mixture was gently mixed and incubated at 30°C while agitating for 30 minutes. The cells were then heat shocked at 42°C 10 for 15 minutes, and the reaction vessel centrifuged in a microfuge at 12,000 rpm for 5-10 seconds, decanted and resuspended into TE (500 μ l, 10 mM Tris-HCl, 1 mM EDTA pH 7.5) followed by recentrifugation. The cells were then diluted into TE (1 ml) and aliquots (200 μ l) were spread onto the selective media previously prepared in 150 mm growth plates (VWR).

15 Alternatively, instead of multiple small reactions, the transformation was performed using a single, large scale reaction, wherein reagent amounts were scaled up accordingly.

The selective media used was a synthetic complete dextrose agar lacking uracil (SCD-Ura) prepared as described in Kaiser et al., *Methods in Yeast Genetics*, Cold Spring Harbor Press, Cold Spring Harbor, NY, p. 208-210 (1994). Transformants were grown at 30°C for 2-3 days.

20 The detection of colonies secreting amylase was performed by including red starch in the selective growth media. Starch was coupled to the red dye (Reactive Red-120, Sigma) as per the procedure described by Biely et al., *Anal. Biochem.*, 172:176-179 (1988). The coupled starch was incorporated into the SCD-Ura agar plates at a final concentration of 0.15% (w/v), and was buffered with potassium phosphate to a pH of 7.0 (50-100 mM final concentration).

25 The positive colonies were picked and streaked across fresh selective media (onto 150 mm plates) in order to obtain well isolated and identifiable single colonies. Well isolated single colonies positive for amylase secretion were detected by direct incorporation of red starch into buffered SCD-Ura agar. Positive colonies were determined by their ability to break down starch resulting in a clear halo around the positive colony visualized directly.

30 4. Isolation of DNA by PCR Amplification

When a positive colony was isolated, a portion of it was picked by a toothpick and diluted into sterile water (30 μ l) in a 96 well plate. At this time, the positive colonies were either frozen and stored for subsequent analysis or immediately amplified. An aliquot of cells (5 μ l) was used as a template for the PCR reaction in a 25 μ l volume containing: 0.5 μ l Klentaq (Clontech, Palo Alto, CA); 4.0 μ l 10 mM dNTP's (Perkin Elmer- 35 Cetus); 2.5 μ l Kentaq buffer (Clontech); 0.25 μ l forward oligo 1; 0.25 μ l reverse oligo 2; 12.5 μ l distilled water.

The sequence of the forward oligonucleotide 1 was:

5'-TGTAAAACGACGGCCAGTTAAATAGACCTGCAATTATTAATCT-3' (SEQ ID NO:16)

The sequence of reverse oligonucleotide 2 was:

5'-CAGGAAACAGCTATGACCACCTGCACACCTGCAAATCCATT-3' (SEQ ID NO:17)

PCR was then performed as follows:

5	a.	Denature	92°C, 5 minutes
	b.	3 cycles of:	Denature 92°C, 30 seconds
			Anneal 59°C, 30 seconds
			Extend 72°C, 60 seconds
10	c.	3 cycles of:	Denature 92°C, 30 seconds
			Anneal 57°C, 30 seconds
			Extend 72°C, 60 seconds
15	d.	25 cycles of:	Denature 92°C, 30 seconds
			Anneal 55°C, 30 seconds
			Extend 72°C, 60 seconds
20	e.	Hold	4°C

The underlined regions of the oligonucleotides annealed to the ADH promoter region and the amylase region, respectively, and amplified a 307 bp region from vector pSST-AMY.0 when no insert was present. Typically, the first 18 nucleotides of the 5' end of these oligonucleotides contained annealing sites for the sequencing primers. Thus, the total product of the PCR reaction from an empty vector was 343 bp. However, signal sequence-fused cDNA resulted in considerably longer nucleotide sequences.

Following the PCR, an aliquot of the reaction (5 μ l) was examined by agarose gel electrophoresis in a 1% agarose gel using a Tris-Borate-EDTA (TBE) buffering system as described by Sambrook et al., supra. Clones resulting in a single strong PCR product larger than 400 bp were further analyzed by DNA sequencing after purification with a 96 Qiaquick PCR clean-up column (Qiagen Inc., Chatsworth, CA).

30

EXAMPLE 3: Isolation of cDNA Clones Encoding Human PRO241

A consensus DNA sequence was assembled relative to other EST sequences as described in Example 1 above. This consensus sequence is herein designated DNA30876. Based on the DNA30876 consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO241.

35 PCR primers (forward and reverse) were synthesized:

forward PCR primer 5'-GGAAATGAGTGCAAACCCCTC-3' (SEQ ID NO:3)

reverse PCR primer 5'-TCCCAAGCTGAACACTCATTCTGC-3' (SEQ ID NO:4)

40 Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA30876 sequence which had the following nucleotide sequence

hybridization probe

5'-GGGTGACGGTGTCCATATCAGAATTGCAGAAGCAAAACTGACCTCAGTT-3' (SEQ ID NO:5)

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO241 gene using the probe oligonucleotide and one of the PCR primers. RNA for construction of the cDNA libraries was isolated from human fetal kidney tissue (LIB29).

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for 5 PRO241 [herein designated as DNA34392-1170] (SEQ ID NO:1) and the derived protein sequence for PRO241.

The entire nucleotide sequence of DNA34392-1170 is shown in Figure 1 (SEQ ID NO:1). Clone DNA34392-1170 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 234-236 and ending at the stop codon at nucleotide positions 1371-1373 (Figure 1). The predicted polypeptide precursor is 379 amino acids long (Figure 2). The full-length PRO241 protein shown in Figure 2 10 has an estimated molecular weight of about 43,302 daltons and a pI of about 7.30. Clone DNA34392-1170 has been deposited with ATCC and is assigned ATCC deposit no. ATCC 209526.

Analysis of the amino acid sequence of the full-length PRO241 polypeptide suggests that it possess significant homology to the various biglycan proteoglycan proteins, thereby indicating that PRO241 is a novel biglycan homolog polypeptide.

EXAMPLE 4: Isolation of cDNA Clones Encoding Human PRO243 by Genomic Walking

Introduction: Human thrombopoietin (THPO) is a glycosylated hormone of 352 amino acids consisting of two domains. The N-terminal domain, sharing 50% similarity to erythropoietin, is responsible for the biological activity. The C-terminal region is required for secretion. The gene for thrombopoietin (THPO) maps to human chromosome 3q27-q28 where the six exons of this gene span 7 kilobase base pairs of genomic DNA (Gurney *et al.*, *Blood* 85: 981-988 (1995)). In order to determine whether there were any genes encoding THPO homologues located in close proximity to THPO, genomic DNA fragments from this region were identified and sequenced. Three P1 clones and one PAC clones (Genome Systems Inc., St. Louis, MO; cat. Nos. P1-2535 and PAC-6539) encompassing the THPO locus were isolated and a 140 kb region was sequenced using the 25 ordered shotgun strategy (Chen *et al.*, *Genomics* 17: 651-656 (1993)), coupled with a PCR-based gap filling approach. Analysis reveals that the region is gene-rich with four additional genes located very close to THPO: tumor necrosis factor-receptor type 1 associated protein 2 (TRAP2) and elongation initiation factor gamma (eIF4g), chloride channel 2 (CLCN2) and RNA polymerase II subunit hRPB17. While no THPO homolog was found in the region, four novel genes have been predicted by computer-assisted gene detection (GRAIL)(Xu *et* 30 *al.*, *Gen. Engin.* 16: 241-253 (1994), the presence of CpG islands (Cross, S. and Bird, A., *Curr. Opin. Genet. & Devel.* 5: 109-314 (1995), and homology to known genes (as detected by WU-BLAST2.0)(Altschul and Gish, *Methods Enzymol.* 266: 460-480 (1996) (<http://blast.wustl.edu/blast/README.html>).

P1 and PAC clones: The initial human P1 clone was isolated from a genomic P1 library (Genome Systems 35 Inc., St. Louis, MO; cat. no.: P1-2535) screened with PCR primers designed from the THPO genomic sequence (A.L. Gurney, *et al.*, *Blood* 85: 981-88 (1995)). PCR primers were designed from the end sequences derived from this P1 clone were then used to screen P1 and PAC libraries (Genome Systems, Cat. Nos.: P1-2535 &

PAC-6539) to identify overlapping clones.

Ordered Shotgun Strategy: The Ordered Shotgun Strategy (OSS) (Chen *et al.*, *Genomics* **17**: 651-656 (1993)) involves the mapping and sequencing of large genomic DNA clones with a hierarchical approach. The P1 or PAC clone was sonicated and the fragments subcloned into lambda vector (λ Bluestar) (Novagen, Inc., 5 Madison, WI; cat. no. 69242-3). The lambda subclone inserts were isolated by long-range PCR (Barnes, W. *Proc. Natl. Acad. Sci. USA* **91**: 2216-2220 (1994) and the ends sequenced. The lambda-end sequences were overlapped to create a partial map of the original clone. Those lambda clones with overlapping end-sequences were identified, the insets subcloned into a plasmid vector (pUC9 or pUC18) and the ends of the plasmid subclones were sequenced and assembled to generate a contiguous sequence. This directed sequencing strategy 10 minimizes the redundancy required while allowing one to scan for and concentrate on interesting regions.

In order to define better the THPO locus and to search for other genes related to the hematopoietin family, four genomic clones were isolated from this region by PCR screening of human P1 and PAC libraries (Genome System, Inc., Cat. Nos.: P1-2535 and PAC-6539). The sizes of the genomic fragments are as follows: P1.t is 40 kb; P1.g is 70 kb; P1.u is 70 kb; and PAC.z is 200 kb. Approximately 80% of the 200 kb genomic DNA region was sequenced by the Ordered Shotgun Strategy (OSS) (Chen *et al.*, *Genomics* **17**: 651-56 (1993), and assembled into contigs using AutoAssemblerTM (Applied Biosystems, Perkin Elmer, Foster City, CA, cat. no. 903227). The preliminary order of these contigs was determined by manual analysis. There were 46 contigs and filling in the gaps was employed. Table 7 summarized the number and sizes of the gaps.

20
25
30
35
Table 7
Summary of the gaps in the 140 kb region

<u>Size of gap</u>	<u>number</u>
< 50 bp	13
50-150 bp	7
150-300 bp	7
300-1000 bp	10
1000-5000 bp	7
> 5000 bp	2 (15,000 bp)

30 *DNA sequencing:* ABI DYE-primerTM chemistry (PE Applied Biosystems, Foster City, CA; Cat. No.: 402112) was used to end-sequence the lambda and plasmid subclones. ABI DYE-terminatorTM chemistry (PE Applied Biosystems, Foster City, CA, Cat. No: 403044) was used to sequence the PCR products with their respective PCR primers. The sequences were collected with an ABI377 instrument. For PCR products larger than 1kb, walking primers were used. The sequences of contigs generated by the OSS strategy in 35 AutoAssemblerTM a (PE Applied Biosystems, Foster City, CA; Cat. No: 903227) and the gap-filling sequencing trace files were imported into SequencherTM (Gene Codes Corp., Ann Arbor, MI) for overlapping and editing.

PCR-Based gap filling Strategy: Primers were designed based on the 5'- and 3'-end sequenced of each contig, avoiding repetitive and low quality sequence regions. All primers were designed to be 19-24-mers with 50-70% G/C content. Oligos were synthesized and gel-purified by standard methods.

Since the orientation and order of the contigs were unknown, permutations of the primers were used in the amplification reactions. Two PCR kits were used: first, XL PCR kit (Perkin Elmer, Norwalk, CT; Cat. No.: N8080205), with extension times of approximately 10 minutes; and second, the Taq polymerase PCR kit (Qiagen Inc., Valencia, CA; Cat. No.: 201223) was used under high stringency conditions if smeared or multiple products were observed with the XL PCR kit. The main PCR product from each successful reactions was extracted from a 0.9% low melting agarose gel and purified with the Geneclean DNA Purification kit prior to sequencing.

10

Analysis: The identification and characterization of coding regions was carried out as follows: First, repetitive sequences were masked using RepeatMasker (A.F.A. Smit & P. Green, http://ftp.genome.washington.edu/RM/RM_details.html) which screens DNA sequences in FastA format against a library of repetitive elements and returns a masked query sequence. Repeats not masked were identified by comparing the sequence to the GenBank database using WUBLAST (Altschul, S & Gish, W., *Methods Enzymol.* **266**: 460-480 (1996) and were masked manually.

Next, known genes were revealed by comparing the genomic regions against Genentech's protein database using the WUBLAST2.0 algorithm and then annotated by aligning the genomic and cDNA sequences for each gene, respectively, using a Needleman-Wunch (Needleman and Wunsch, *J. Mol. Biol.* **48**: 443-453 (1970) algorithm to find regions of local identity between sequences which are otherwise largely dissimilar. The strategy results in detection of all exons of the five known genes in the region, THPO, TRAP2, eIF4g, CLCN2 and hRPB17 (Table 8).

Table 8

Summary of known genes located in the 140 kb region analyzed

<u>Known genes</u>	<u>Map position</u>
eukaryotic translation initiation factor 4 gamma	3q27-qter
thrombopoietin	3q26-q27
chloride channel 2	3q26-qter
TNF receptor associated protein 2	not previously mapped
RNA polymerase II subunit hRPB17	not previously mapped

Finally, novel transcription units were predicted using a number of approaches. CpG islands (S. Cross & Bird, A., *Curr. Opin. Genet. Dev.* **5**: 109-314 (1995) islands were used to define promoter regions and were identified as clusters of sites cleaved by enzymes recognizing GC-rich, 6 or 8-mer palindromic sequences. CpG islands are usually associated with promoter regions of genes. WUBLAST2.0 analysis of short genomic regions (10-20 kb) versus GenBank revealed matches to ESTs. The individual EST sequences (or where possible, their

sequence chromatogram files) were retrieved and assembled with Sequencher to provide a theoretical cDNA sequence (designated herein as DNA34415). GRAIL2 (ApoCom Inc., Knoxville, TN, command line version for the DEC alpha) was used to predict a novel exon. The five known genes in the region served as internal controls for the success of the GRAIL algorithm.

5 *Isolation:* Chordin cDNA clones were isolated from an oligo-dT-primed human fetal lung library. Human fetal lung polyA⁺ RNA was purchased from Clontech (cat #6528-1, lot #43777) and 5 mg used to construct a cDNA library in pKR5B (Genentech, LIB26). The 3'-primer (pGACTAGTTCTAGATCGCAGCGGCCCTTTTTTTTT) (SEQ ID NO:8) and the 5'-linker (pCGGACGCGTGGGCCTGCGACCCAGCT) (SEQ ID NO:9) were designed to introduce SalI and NotI 10 restriction sites. Clones were screened with oligonucleotide probes designed from the putative human chordin cDNA sequence (DNA34415) deduced by manually "splicing" together the proposed genomic exons of the gene. PCR primers flanking the probes were used to confirm the identity of the cDNA clones prior to sequencing.

The screening oligonucleotides probes were the following:

OL15640 34415.p1 5'-GCCGCTCCCCGAACGGGCAGCGGCTCCTCTCAGAA-3' (SEQ ID NO:10) and
OL15642 34415.p2 5'-GGCGCACAGCACGCAGCGCATCACCCGAATGGCTC-3' (SEQ ID NO:11); and
the flanking probes used were the following:
OL15639 34415.f1 5'-GTGCTGCCATCCGTTCTGAGAAGGA-3' (SEQ ID NO:12) and
OL15643 34415.r 5'-GCAGGGTGCTCAAACAGGACAC-3' (SEQ ID NO:13).

20 **EXAMPLE 5: Northern Blot and *in situ* RNA Hybridization Analysis of PRO243**

Expression of PRO243 mRNA in human tissues was examined by Northern blot analysis. Human polyA⁺ RNA blots derived from human fetal and adult tissues (Clontech, Palo Alto, CA; Cat. Nos. 7760-1 and 7756-1) were hybridized to a ³²P-labelled cDNA fragments probe based on the full length PRO243 cDNA. Blots were incubated with the probes in hybridization buffer (5X SSPE; 2X Denhardt's solution; 100 mg/mL denatured 25 sheared salmon sperm DNA; 50% formamide; 2% SDS) for 60 hours at 42°C. The blots were washed several times in 2X SSC; 0.05% SDS for 1 hour at room temperature, followed by a high stringency wash 30 minute wash in 0.1X SSC; 0.1% SDS at 50°C and autoradiographed. The blots were developed after overnight exposure by phosphorimager analysis (Fuji).

PRO243 mRNA transcripts were detected. Analysis of the expression pattern showed the strongest 30 signal of the expected 4.0 kb transcript in adult and fetal liver and a very faint signal in the adult kidney. Fetal brain, lung and kidney were negative, as were adult heart, brain, lung and pancreas. Smaller transcripts were observed in placenta (2.0 kb), adult skeletal muscle (1.8 kb) and fetal liver (2.0 kb).

In situ hybridization of adult human tissue of PRO243 gave a positive signal in the cleavage line of the developing synovial joint forming between the femoral head and acetabulum. All other tissues were negative. 35 Additional sections of human fetal face, head, limbs and mouse embryos were examined. Expression in human fetal tissues was observed adjacent to developing limb and facial bones in the perosteal mesenchyme. The expression was highly specific and was often adjacent to areas undergoing vascularization. Expression was also

observed in the developing temporal and occipital lobes of the fetal brain, but was not observed elsewhere in the brain. In addition, expression was seen in the ganglia of the developing inner ear. No expression was seen in any of the mouse tissues with the human probes.

5 *In situ* hybridization was performed using an optimized protocol, using PCR-generating ^{33}P -labeled riboprobes. (Lu and Gillett, *Cell Vision* 1: 169-176 (1994)). Formalin-fixed, paraffin-embedded human fetal and adult tissues were sectioned, deparaffinized, deproteinated in proteinase K (20 g/ml) for 15 minutes at 37°C, and further processed for *in situ* hybridization as described by Lu and Gillett (1994). A ^{33}P -UTP-labeled antisense riboprobe was generated from a PCR product and hybridized at 55°C overnight. The slides were dipped in Kodak NTB2 nuclear track emulsion and exposed for 4 weeks.

10 EXAMPLE 6: Isolation of cDNA clones Encoding Human PRO299

A cDNA sequence designated herein as DNA28847 (Figure 7; SEQ ID NO:18) was isolated as described in Example 2 above. After further analysis, a 3' truncated version of DNA28847 was found and is herein designated DNA35877 (Figure 8; SEQ ID NO:19). Based on the DNA35877 sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO299. Forward and reverse PCR primers generally range from 20 to 30 nucleotides and are often designed to give a PCR product of about 100-1000 bp in length. The probe sequences are typically 40-55 bp in length. In some cases, additional oligonucleotides are synthesized when the consensus sequence is greater than about 1-1.5kbp. In order to screen several libraries for a full-length clone, DNA from the libraries was screened by PCR amplification, as per Ausubel et al., Current Protocols in Molecular Biology, with the PCR primer pair. A positive library was then used to isolate clones encoding the gene of interest using the probe oligonucleotide and one of the primer pairs.

Forward and reverse PCR primers were synthesized:

forward PCR primer 5'-CTCTGGAAGGTCACGGCACAGG-3' (SEQ ID NO:20)

reverse PCR primer 5'-CTCAGTCGGTGGCAAAGCTCTC-3' (SEQ ID NO:21)

25 Additionally, a synthetic oligonucleotide hybridization probe was constructed from the DNA35877 sequence which had the following nucleotide sequence

hybridization probe

5'-CAGTGCTCCCTCATAGATGGACGAAAGTGTGACCCCCCTTCAGGCGAGAGCTTGCCAACCG
AACTGA-3' (SEQ ID NO:22)

30 In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with one of the PCR primer pairs identified above. A positive library was then used to isolate clones encoding the PRO299 sequence using the probe oligonucleotide.

RNA for construction of the cDNA libraries was isolated from human fetal brain tissue. The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRKB or pRKD;

pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., *Science*, 253:1278-1280 (1991) in the unique XhoI and NotI sites.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO299 [herein designated as DNA39976-1215] (SEQ ID NO:14) and the derived protein sequence for PRO299.

5 The entire nucleotide sequence of DNA39976-1215 is shown in Figure 5 (SEQ ID NO:14). Clone DNA39976-1215 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 111-113 and ending at the stop codon at nucleotide positions 2322-2324 (Figure 5). The predicted polypeptide precursor is 737 amino acids long (Figure 6). Important regions of the polypeptide sequence encoded by clone DNA39976-1215 have been identified and include the following: a signal peptide 10 corresponding to amino acids 1-28, a putative transmembrane region corresponding to amino acids 638-662, 10 EGF repeats, corresponding to amino acids 80-106, 121-203, 336-360, 378-415, 416-441, 454-490, 491-528, 529-548, 567-604, and 605-622, respectively, and 10 potential N-glycosylation sites, corresponding to amino acids 107-120, 204-207, 208-222, 223-285, 286-304, 361-374, 375-377, 442-453, 549-563, and 564-566, respectively. Clone DNA39976-1215 has been deposited with ATCC and is assigned ATCC deposit no. ATCC 209524.

15 Analysis of the amino acid sequence of the full-length PRO299 polypeptide suggests that portions of it possess significant homology to the notch protein, thereby indicating that PRO299 may be a novel notch protein homolog and have activity typical of the notch protein.

20 EXAMPLE 7: Isolation of cDNA Clones Encoding Human PRO323

A consensus DNA sequence was assembled relative to other EST sequences as described in Example 1 above. This consensus sequence is herein designated DNA30875. Based on the DNA30875 consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO323.

25 PCR primers (two forward and one reverse) were synthesized:

forward PCR primer 1 5'-AGTTCTGGTCAGCCTATGTGCC-3' (SEQ ID NO:25)

forward PCR primer 2 5'-CGTGATGGTGTCTTGTCCATGGG-3' (SEQ ID NO:26)

reverse PCR primer 5'-CTCCACCAATCCCGATGAACCTGG-3' (SEQ ID NO:27)

30 Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA30875 sequence which had the following nucleotide sequence

hybridization probe

5'-GAGCAGATTGACCTCATACGCCGCATGTGTGCCTCCTATTCTGAGCTGGA-3' (SEQ ID NO:28)

35 In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pairs identified above. A positive library was then used to isolate clones encoding the PRO323 gene using the probe oligonucleotide and one of the PCR primers. RNA for construction of the cDNA libraries was isolated from human fetal liver tissue (LIB6).

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO323 [herein designated as DNA35595-1228] (SEQ ID NO:23) and the derived protein sequence for PRO323.

The entire nucleotide sequence of DNA35595-1228 is shown in Figure 9 (SEQ ID NO:23). Clone DNA35595-1228 contains a single open reading frame with an apparent translational initiation site at nucleotide 5 positions 110-112 and ending at the stop codon at nucleotide positions 1409-1411 (Figure 9). The predicted polypeptide precursor is 433 amino acids long (Figure 10). The full-length PRO323 protein shown in Figure 10 has an estimated molecular weight of about 47,787 daltons and a pI of about 6.11. Clone DNA35595-1228 has been deposited with ATCC and is assigned ATCC deposit no. 209528.

Analysis of the amino acid sequence of the full-length PRO323 polypeptide suggests that portions of 10 it possess significant homology to various dipeptidase proteins, thereby indicating that PRO323 may be a novel dipeptidase protein.

EXAMPLE 8: Isolation of cDNA Clones Encoding Human PRO327

An expressed sequence tag (EST) DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA) was searched and various EST sequences were identified which showed certain degrees of homology to human prolactin receptor protein. Those EST sequences were aligned using phrap and a consensus sequence was obtained. This consensus DNA sequence was then extended using repeated cycles of BLAST and phrap to extend the consensus sequence as far as possible using the sources of EST sequences discussed above. The extended assembly sequence is herein designated DNA38110. The above searches were performed using the computer program BLAST or BLAST2 (Altshul et al., *Methods in Enzymology* 266:460-480 (1996)). Those comparisons resulting in a BLAST score of 70 (or in some cases 90) or greater that did not encode known 20 proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, Washington).

Based upon the DNA38110 consensus sequence obtained as described above, oligonucleotides were 25 synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO327.

PCR primers (forward and reverse) were synthesized as follows:

forward PCR primer 5'-CCCGCCCGACGTGCACGTGAGCC-3' (SEQ ID NO:33)

reverse PCR primer 5'-TGAGCCAGCCCAGGAAGTGTGCTTG-3' (SEQ ID NO:34)

30 Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA38110 consensus sequence which had the following nucleotide sequence

hybridization probe

5'-CAAGTGCCTGCAACCCCTTGGCATCTATGGCTCAAGAAAGCCGGAT-3' (SEQ ID NO:35)

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was 35 screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO327 gene using the probe oligonucleotide and one of the PCR primers. RNA for construction of the cDNA libraries was isolated from human fetal lung tissue (LIB26).

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO327 [herein designated as DNA38113-1230] (SEQ ID NO:16) and the derived protein sequence for PRO327.

The entire nucleotide sequence of DNA38113-1230 is shown in Figure 13 (SEQ ID NO:31). Clone DNA38113-1230 contains a single open reading frame with an apparent translational initiation site at nucleotide 5 positions 119-121 and ending at the stop codon at nucleotide positions 1385-1387 (Figure 13). The predicted polypeptide precursor is 422 amino acids long (Figure 14). The full-length PRO327 protein shown in Figure 14 has an estimated molecular weight of about 46,302 daltons and a pI of about 9.42. Clone DNA38113-1230 has been deposited with ATCC and is assigned ATCC deposit no. ATCC 209530.

Analysis of the amino acid sequence of the full-length PRO327 polypeptide suggests that it possess 10 significant homology to the human prolactin receptor protein, thereby indicating that PRO327 may be a novel prolactin binding protein.

EXAMPLE 9: Isolation of cDNA Clones Encoding Human PRO233

A consensus DNA sequence was assembled relative to other EST sequences as described in Example 15 above. This consensus sequence is herein designated DNA30945. Based on the DNA30945 consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO233.

PCR primers were synthesized as followed:

forward PCR primer 5'-GGTGAAGGCAGAAATTGGAGATG-3' (SEQ ID NO:38)

reverse PCR primer 5'-ATCCCAGTCATCAGCCTGTTACC-3' (SEQ ID NO:39)

Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA30945 sequence which had the following nucleotide sequence

hybridization probe

5'-GCTGGTAGTCTATACATCAGATTGTTGCTACACAAGATCCTCAG-3'

25 (SEQ ID NO:40)

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO233 gene using the probe oligonucleotide. RNA for construction of the cDNA libraries was isolated from human fetal brain tissue.

30 DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO233 [herein designated as DNA34436-1238] (SEQ ID NO:36) and the derived protein sequence for PRO233.

The entire nucleotide sequence of DNA34436-1238 is shown in Figure 15 (SEQ ID NO:36). Clone 35 DNA34436-1238 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 101-103 and ending at the stop codon at nucleotide positions 1001-1003 (Figure 15). The predicted polypeptide precursor is 300 amino acids long (Figure 16). The full-length PRO233 protein shown in Figure 16 has an estimated molecular weight of about 32,964 daltons and a pI of about 9.52. In addition, regions of

interest including the signal peptide and a putative oxidoreductase active site, are designated in Figure 16. Clone DNA34436-1238 has been deposited with ATCC and is assigned ATCC deposit no. ATCC 209523

Analysis of the amino acid sequence of the full-length PRO233 polypeptide suggests that portions of it possess significant homology to various reductase proteins, thereby indicating that PRO233 may be a novel reductase.

5

EXAMPLE 10: Isolation of cDNA Clones Encoding Human PRO344

A consensus DNA sequence was assembled relative to other EST sequences as described in Example 1 above. This consensus sequence is herein designated DNA34398. Based on the DNA34398 consensus sequences, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence 10 of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO344.

Based on the DNA34398 consensus sequence, forward and reverse PCR primers were synthesized as follows:

<u>forward PCR primer</u>	(34398.f1)	5'-TACAGGCCAGTCAGGACCAGGGG-3'	(SEQ ID NO:43)
<u>forward PCR primer</u>	(34398.f2)	5'-AGCCAGCCTCGCTCTCGG-3'	(SEQ ID NO:44)
<u>forward PCR primer</u>	(34398.f3)	5'-GTCTGCGATCAGGTCTGG-3'	(SEQ ID NO:45)
<u>reverse PCR primer</u>	(34398.r1)	5'-GAAAGAGGCAATGGATTGCG-3'	(SEQ ID NO:46)
<u>reverse PCR primer</u>	(34398.r2)	5'-GACTTACACTTGCCAGCACAGCAC-3'	(SEQ ID NO:47)

Additionally, a synthetic oligonucleotide hybridization probe was constructed from the DNA34398 consensus sequence which had the following nucleotide sequence

<u>hybridization probe</u>	(34398.p1)	5'-GGAGCACCAACTGGAGGGTCCGGAGTAGCGAGCGCCCCGAAG-3'	(SEQ ID NO:48)
----------------------------	------------	--	----------------

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with one of the PCR primer pairs identified above. A positive library was then used to isolate clones encoding the PRO344 genes using the probe oligonucleotide and one of the PCR primers.

25 RNA for construction of the cDNA libraries was isolated from human fetal kidney tissue.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO344 [herein designated as DNA40592-1242] (SEQ ID NO:41) and the derived protein sequence for PRO344.

The entire nucleotide sequence of DNA40592-1242 is shown in Figure 17 (SEQ ID NO:41). Clone

30 DNA40592-1242 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 227-229 and ending at the stop codon at nucleotide positions 956-958 (Figure 17). The predicted polypeptide precursor is 243 amino acids long (Figure 18). Important regions of the native PRO344 amino acid sequence include the signal peptide, the start of the mature protein, and two potential N-myristoylation sites as shown in Figure 18. Clone DNA40592-1242 has been deposited with the ATCC and is assigned ATCC deposit 35 no. ATCC 209492

Analysis of the amino acid sequence of the full-length PRO344 polypeptides suggests that portions of them possess significant homology to various human and murine complement proteins, thereby indicating that

PRO344 may be a novel complement protein.

EXAMPLE 11: Isolation of cDNA Clones Encoding Human PRO347

A consensus DNA sequence was assembled relative to other EST sequences as described in Example 1 above. This consensus sequence is herein designated DNA39499. Based on the DNA39499 consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO347.

PCR primers (forward and reverse) were synthesized as follows:

forward PCR primer 5'-AGGAACCTCTGGATCGGGCTCAC-3' (SEQ ID NO:51)

reverse PCR primer 5'-GGGTCTGGGCCAGGTGGAAGAGAG-3' (SEQ ID NO:52)

10 Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA39499 sequence which had the following nucleotide sequence

hybridization probe

5'-GCCAAGGACTCCTCCGCTGGGCCACAGGGGAGCACCCAGGCCTTC-3' (SEQ ID NO:53)

15 In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO347 gene using the probe oligonucleotide and one of the PCR primers. RNA for construction of the cDNA libraries was isolated from human fetal kidney tissue (LIB228).

20 DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO347 [herein designated as DNA44176-1244] (SEQ ID NO:49) and the derived protein sequence for PRO347.

25 The entire nucleotide sequence of DNA44176-1244 is shown in Figure 19 (SEQ ID NO:49). Clone DNA44176-1244 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 123-125 and ending at the stop codon at nucleotide positions 1488-1490 (Figure 19). The predicted polypeptide precursor is 455 amino acids long (Figure 20). The full-length PRO347 protein shown in Figure 20 has an estimated molecular weight of about 50,478 daltons and a pI of about 8.44. Clone DNA44176-1244 has been deposited with ATCC and is assigned ATCC deposit no. ATCC 209532

30 Analysis of the amino acid sequence of the full-length PRO347 polypeptide suggests that portions of it possess significant homology to various cysteine-rich secretory proteins, thereby indicating that PRO347 may be a novel cysteine-rich secretory protein.

30

EXAMPLE 12: Isolation of cDNA Clones Encoding Human PRO354

An expressed sequence tag (EST) DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA) was searched and various EST sequences were identified which possessed certain degrees of homology with the inter-alpha-trypsin inhibitor heavy chain and with one another. Those homologous EST sequences were then 35 aligned and a consensus sequence was obtained. The obtained consensus DNA sequence was then extended using repeated cycles of BLAST and phrap to extend the consensus sequence as far as possible using homologous EST sequences derived from both public EST databases (e.g., GenBank) and a proprietary EST DNA database

(LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The extended assembly sequence is herein designated DNA39633. The above searches were performed using the computer program BLAST or BLAST2 (Altshul et al., *Methods in Enzymology* 266:460-480 (1996)). Those comparisons resulting in a BLAST score of 70 (or in some cases 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, Washington).

5 Based on the DNA39633 consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO354. Forward and reverse PCR primers generally range from 20 to 30 nucleotides and are often designed to give a PCR product of about 100-1000 bp in length. The probe sequences are typically 40-55 bp in length. In some cases, additional oligonucleotides are synthesized when the consensus 10 sequence is greater than about 1-1.5 kbp. In order to screen several libraries for a full-length clone, DNA from the libraries was screened by PCR amplification, as per Ausubel et al., *Current Protocols in Molecular Biology*, with the PCR primer pair. A positive library was then used to isolate clones encoding the gene of interest using the probe oligonucleotide and one of the primer pairs.

PCR primers were synthesized as follows:

15 forward PCR primer 1 (39633.f1) 5'-GTGGGAACCAAACCTCCGGCAGACC-3' (SEQ ID NO:56)

forward PCR primer 2 (39633.f2) 5'-CACATCGAGCGTCTCTGG-3' (SEQ ID NO:57)

reverse PCR primer (39633.r1) 5'-AGCCGCTCCTCTCCGGTTCATCG-3' (SEQ ID NO:58)

Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA39633 sequence which had the following nucleotide sequence

20 hybridization probe

5'-TGGAAGGACCACTTGATATCAGTCACTCCAGACAGCATCAGGGATGGG-3' (SEQ ID NO:59)

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pairs identified above. A positive library was then used to isolate clones encoding the PRO354 gene using the probe oligonucleotide and one of the PCR primers.

25 RNA for construction of the cDNA libraries was isolated from human fetal kidney tissue (LIB227).

The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to Sall hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRKB or 30 pRKD; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., *Science*, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO354 [herein designated as DNA44192-1246] (SEQ ID NO:54) and the derived protein sequence for PRO354.

35 The entire nucleotide sequence of DNA44192-1246 is shown in Figure 21 (SEQ ID NO:54). Clone DNA44192-1246 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 72-74 and ending at the stop codon at nucleotide positions 2154-2156 (Figure 21). The predicted

polypeptide precursor is 694 amino acids long (Figure 22). The full-length PRO354 protein shown in Figure 22 has an estimated molecular weight of about 77,400 daltons and a pI of about 9.54. Clone DNA44192-1246 has been deposited with ATCC and is assigned ATCC deposit no. ATCC 209531.

Analysis of the amino acid sequence of the full-length PRO354 polypeptide suggests that it possess significant homology to the inter-alpha-trypsin inhibitor heavy chain protein, thereby indicating that PRO354 may 5 be a novel inter-alpha-trypsin inhibitor heavy chain protein homolog.

EXAMPLE 13: Isolation of cDNA Clones Encoding Human PRO355

A consensus DNA sequence was assembled relative to other EST sequences using BLAST and phrap as described in Example 1 above. This consensus sequence is herein designated DNA35702. Based on the 10 DNA35702 consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO355.

Forward and reverse PCR primers were synthesized as follows:

<u>forward PCR primer</u>	5'-GGCTTCTGCTGTTGCTCTTCTCCG-3'	(SEQ ID NO:62)
<u>forward PCR primer</u>	5'-GTACACTGTGACCAGTCAGC-3'	(SEQ ID NO:63)
<u>forward PCR primer</u>	5'-ATCATCACAGATTCCCAGC-3'	(SEQ ID NO:64)
<u>reverse PCR primer</u>	5'-TTCAATCTCCTCACCTCCACCGC-3'	(SEQ ID NO:65)
<u>reverse PCR primer</u>	5'-ATAGCTGTGTCTGCGTCTGCG-3'	(SEQ ID NO:66)

Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA35702 sequence which had the following nucleotide sequence:

hybridization probe

5'-CGCGGCACTGATCCCCACAGGTGATGGGCAGAATCTGTTACGAAAGACG-3' (SEQ ID NO:67)

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with one of the PCR primer pairs identified above. A positive library was then 20 used to isolate clones encoding the PRO355 gene using the probe oligonucleotide. RNA for construction of the cDNA libraries was isolated from human fetal liver tissue.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO355 [herein designated as DNA39518-1247] (SEQ ID NO:60) and the derived protein sequence for PRO355.

30 The entire nucleotide sequence of DNA39518-1247 is shown in Figure 23 (SEQ ID NO:60). Clone DNA39518-1247 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 22-24 and ending at the stop codon at nucleotide positions 1342-1344 (Figure 23). The predicted polypeptide precursor is 440 amino acids long (Figure 24). The full-length PRO355 protein shown in Figure 24 has an estimated molecular weight of about 48,240 daltons and a pI of about 4.93. In addition, regions of 35 interest including the signal peptide, Ig repeats in the extracellular domain, potential N-glycosylation sites, and the potential transmembrane domain, are designated in Figure 24. Clone DNA39518-1247 has been deposited with ATCC and is assigned ATCC deposit no. ATCC 209529.

Analysis of the amino acid sequence of the full-length PRO355 polypeptide suggests that portions of it possess significant homology to the CRTAM protein, thereby indicating that PRO355 may be CRTAM protein.

EXAMPLE 14: Isolation of cDNA Clones Encoding Human PRO357

The sequence expression tag clone no. "2452972" by Incyte Pharmaceuticals, Palo Alto, CA was used 5 to begin a data base search. The extracellular domain (ECD) sequences (including the secretion signal, if any) of from about 950 known secreted proteins from the Swiss-Prot public protein database were used to search expressed sequence tag (EST) databases which overlapped with a portion of Incyte EST clone no. "2452972". The EST databases included public EST databases (e.g., GenBank) and a proprietary EST DNA database 10 (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST2 (Altshul et al., *Methods in Enzymology* 266:460-480 (1996)) as a comparison of the ECD protein sequences to a 6 frame translation of the EST sequence. Those comparisons resulting in a BLAST score of 70 (or in some cases 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, Washington).

A consensus DNA sequence was then assembled relative to other EST sequences using phrap. This 15 consensus sequence is herein designated DNA37162. In this case, the consensus DNA sequence was extended using repeated cycles of BLAST and phrap to extend the consensus sequence as far as possible using the sources of EST sequences discussed above.

Based on the DNA37162 consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR 20 a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO357. Forward and reverse PCR primers generally range from 20 to 30 nucleotides and are often designed to give a PCR product of about 100-1000 bp in length. The probe sequences are typically 40-55 bp in length. In some cases, additional oligonucleotides are synthesized when the consensus 25 sequence is greater than about 1-1.5kbp. In order to screen several libraries for a full-length clone, DNA from the libraries was screened by PCR amplification, as ber Ausubel et al., *Current Protocols in Molecular Biology*, with the PCR primer pair. A positive library was then used to isolate clones encoding the gene of interest using the probe oligonucleotide and one of the primer pairs.

PCR primers were synthesized as follows:

forward primer 1: 5'-CCCTCCACTGCCCAACCGACTG-3' (SEQ ID NO:70);
30 reverse primer 1: 5'-CGGTTCTGGGGACGTTAGGGCTCG-3' (SEQ ID NO:71); and
forward primer 2: 5'-CTGCCAACCGTCCACCTGCCTCAAT-3' (SEQ ID NO:72).
Additionally, two synthetic oligonucleotidehybridizationprobes were constructed from the consensus DNA37162 sequence which had the following nucleotide sequences:
hybridization probe 1:
35 5'-AGGACTGCCACCGTCCACCTGCCTCAATGGGGCACATGCCACC-3' (SEQ ID NO:73); and
hybridization probe 2:
5'-ACGAAAGCCCTACATCTAAGCCAGAGAGACAGGGCAGCTGGG-3' (SEQ ID NO:74).

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with a PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO357 gene using the probe oligonucleotide and one of the PCR primers.

RNA for construction of the cDNA libraries was isolated from human fetal liver tissue. The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRK8 or pRKD; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., *Science*, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

10 DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO357 [herein designated as DNA44804-1248] (SEQ ID NO:68) and the derived protein sequence for PRO357.

15 The entire nucleotide sequence of DNA44804-1248 is shown in Figure 25 (SEQ ID NO:68). Clone DNA44804-1248 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 137-139 and ending at the stop codon at nucleotide positions 1931-1933 (Figure 25). The predicted polypeptide precursor is 598 amino acids long (Figure 26). Clone DNA44804-1248 has been deposited with ATCC and is assigned ATCC deposit no. ATCC 209527.

20 Analysis of the amino acid sequence of the full-length PRO357 polypeptide therefore suggests that portions of it possess significant homology to ALS, thereby indicating that PRO357 may be a novel leucine rich repeat protein related to ALS.

EXAMPLE 15: Isolation of cDNA Clones Encoding Human PRO715

25 A proprietary EST DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA) was searched for EST sequences encoding polypeptides having homology to human TNF- α . This search resulted in the identification of Incyte Expressed Sequence Tag No. 2099855.

A consensus DNA sequence was then assembled relative to other EST sequences using seqext and "phrap" (Phil Green, University of Washington, Seattle, Washington). This consensus sequence is herein designated DNA52092. Based upon the alignment of the various EST clones identified in this assembly, a single EST clone from the Merck/Washington University EST set (EST clone no. 725887, Accession No. AA292358) 30 was obtained and its insert sequenced. The full-length DNA52722-1229 sequence was then obtained from sequencing the insert DNA from EST clone no. 725887.

The entire nucleotide sequence of DNA52722-1229 is shown in Figure 27 (SEQ ID NO:75). Clone DNA52722-1229 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 114-116 and ending at the stop codon at nucleotide positions 864-866 (Figure 27). The predicted polypeptide is 250 amino acids long (Figure 28). The full-length PRO715 protein shown in Figure 28 has an estimated molecular weight of about 27,433 daltons and a pI of about 9.85.

Analysis of the amino acid sequence of the full-length PRO715 polypeptide suggests that it possesses significant homology to members of the tumor necrosis factor family of proteins, thereby indicating that PRO715 is a novel tumor necrosis factor protein.

EXAMPLE 16: Isolation of cDNA Clones Encoding Human PRO353

5 A consensus DNA sequence was assembled relative to other EST sequences using phrap as described in Example 1 above. This consensus sequences is herein designated DNA36363. The consensus DNA sequence was extended using repeated cycles of BLAST and phrap to extend the consensus sequence as far as possible using the sources of EST sequences discussed above. Based on the DNA36363 consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, 10 and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO353.

Based on the DNA36363 consensus sequence, forward and reverse PCR primers were synthesized as follows:

forward PCR primer 5'-TACAGGCCAGTCAGGACCAGGGG-3' (SEQ ID NO:79)

reverse PCR primer 5'-CTGAAGAAGTAGAGGCCGGCACG-3' (SEQ ID NO:80).

15 Additionally, a synthetic oligonucleotide hybridization probe was constructed from the DNA36363 consensus sequence which had the following nucleotide sequence:

hybridization probe

5'-CCCGGTGCTTGCCTGCTGTGACCCCGGTACCTCCATGTACCCGG-3' (SEQ ID NO:81)

20 In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with one of the PCR primer pairs identified above. A positive library was then used to isolate clones encoding the PRO353 gene using the probe oligonucleotide and one of the PCR primers. RNA for construction of the cDNA libraries was isolated from human fetal kidney tissue.

25 DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO353 [herein designated as DNA41234-1242] (SEQ ID NO:77) and the derived protein sequence for PRO353.

The entire nucleotide sequence of DNA41234-1242 is shown in Figure 29 (SEQ ID NO:77). Clone DNA41234-1242 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 305-307 and ending at the stop codon at nucleotide positions 1148-1150 (Figure 29). The predicted polypeptide precursor is 281 amino acids long (Figure 30). Important regions of the amino acid sequence 30 encoded by PRO353 include the signal peptide, corresponding to amino acids 1-26, the start of the mature protein at amino acid position 27, a potential N-glycosylation site, corresponding to amino acids 93-98 and a region which has homology to a 30 kd adipocyte complement-related protein precursor, corresponding to amino acids 99-281. Clone DNA41234-1242 has been deposited with the ATCC and is assigned ATCC deposit no. ATCC 209618.

35 Analysis of the amino acid sequence of the full-length PRO353 polypeptides suggests that portions of them possess significant homology to portions of human and murine complement proteins, thereby indicating that PRO353 may be a novel complement protein.

EXAMPLE 17: Isolation of cDNA Clones Encoding Human PRO361

A consensus DNA sequence was assembled relative to other EST sequences using phrap as described in Example 1 above. This consensus sequence is herein designated DNA40654. Based on the DNA40654 consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO361.

Forward and reverse PCR primers were synthesized as follows:

<u>forward PCR primer</u>	5'-AGGGAGGATTATCCTTGACCTTGAAAGACC-3'	(SEQ ID NO:84)
<u>forward PCR primer</u>	5'-GAAGCAAGTGCCAGCTC-3'	(SEQ ID NO:85)
<u>forward PCR primer</u>	5'-CGGGTCCCTGCTCTTGG-3'	(SEQ ID NO:86)
10 <u>reverse PCR primer</u>	5'-CACCGTAGCTGGAGCGCACTCAC-3'	(SEQ ID NO:87)
<u>reverse PCR primer</u>	5'-AGTGTAAAGTCAGCTCCC-3'	(SEQ ID NO:88)

Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA40654 sequence which had the following nucleotide sequence

hybridization probe

5'- GCTTCCTGACACTAAGGCTGTCTGCTAGTCAGAATTGCCTAAAAAGAG-3' (SEQ ID NO:89)

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with one of the PCR primer pairs identified above. A positive library was then used to isolate clones encoding the PRO361 gene using the probe oligonucleotide. RNA for construction of the cDNA libraries was isolated from human fetal kidney tissue.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO361 [herein designated as DNA45410-1250] (SEQ ID NO:82) and the derived protein sequence for PRO361.

The entire nucleotide sequence of DNA45410-1250 is shown in Figure 31 (SEQ ID NO:82). Clone DNA45410-1250 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 226-228 and ending at the stop codon at nucleotide positions 1519-1521 (Figure 31). The predicted polypeptide precursor is 431 amino acids long (Figure 32). The full-length PRO361 protein shown in Figure 32 has an estimated molecular weight of about 46,810 daltons and a pI of about 6.45. In addition, regions of interest including the transmembrane domain (amino acids 380-409) and sequences typical of the arginase family of proteins (amino acids 3-14 and 39-57) are designated in Figure 32. Clone DNA45410-1250 has been deposited with ATCC and is assigned ATCC deposit no. ATCC 209621.

Analysis of the amino acid sequence of the full-length PRO361 polypeptide suggests that portions of it possess significant homology to the mucin and/or chitinase proteins, thereby indicating that PRO361 may be a novel mucin and/or chitinase protein.

35 EXAMPLE 18: Isolation of cDNA Clones Encoding Human PRO365

A consensus DNA sequence was assembled relative to other EST sequences using phrap as described in Example 1 above. This consensus sequence is herein designated DNA35613. Based on the DNA35613

consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO365.

Forward and reverse PCR primers were synthesized as follows:

<u>forward PCR primer</u>	5'-AATGTGACCACTGGACTCCC-3'	(SEQ ID NO:92)
5 <u>forward PCR primer</u>	5'-AGGCTTGGAACTCCCTTC-3'	(SEQ ID NO:93)
<u>reverse PCR primer</u>	5'-AAGATTCTTGAGCGATTCCAGCTG-3'	(SEQ ID NO:94)

Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA35613 sequence which had the following nucleotide sequence

hybridization probe

10 5'-AATCCCTGCTCTTCATGGTACCTATGACGACGGAAGCACAAAGACTG-3' (SEQ ID NO:95)

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with one of the PCR primer pairs identified above. A positive library was then used to isolate clones encoding the PRO365 gene using the probe oligonucleotide and one of the PCR primers. RNA for construction of the cDNA libraries was isolated from human fetal kidney tissue.

15 DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO365 [herein designated as DNA46777-1253] (SEQ ID NO:90) and the derived protein sequence for PRO365.

20 The entire nucleotide sequence of DNA46777-1253 is shown in Figure 33 (SEQ ID NO:90). Clone DNA46777-1253 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 15-17 and ending at the stop codon at nucleotide positions 720-722 (Figure 33). The predicted polypeptide precursor is 235 amino acids long (Figure 34). Important regions of the polypeptide sequence encoded by clone DNA46777-1253 have been identified and include the following: a signal peptide corresponding to amino acids 1-20, the start of the mature protein corresponding to amino acid 21, and multiple potential N-glycosylation sites as shown in Figure 34. Clone DNA46777-1253 has been deposited with ATCC 25 and is assigned ATCC deposit no. ATCC 209619.

Analysis of the amino acid sequence of the full-length PRO365 polypeptide suggests that portions of it possess significant homology to the human 2-19 protein, thereby indicating that PRO365 may be a novel human 2-19 protein homolog.

30 EXAMPLE 19: Use of PRO Polypeptide-Encoding Nucleic Acid as Hybridization Probes

The following method describes use of a nucleotide sequence encoding PRO as a hybridization probe.

DNA comprising the coding sequence of full-length or mature PRO as disclosed herein is employed as a probe to screen for homologous DNAs (such as those encoding naturally-occurring variants of PRO) in human tissue cDNA libraries or human tissue genomic libraries.

35 Hybridization and washing of filters containing either library DNAs is performed under the following high stringency conditions. Hybridization of radiolabeled PRO-derived probe to the filters is performed in a solution of 50% formamide, 5x SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 50 mM sodium phosphate, pH

6.8, 2x Denhardt's solution, and 10% dextran sulfate at 42°C for 20 hours. Washing of the filters is performed in an aqueous solution of 0.1x SSC and 0.1% SDS at 42°C.

DNAs having a desired sequence identity with the DNA encoding full-length native sequence PRO can then be identified using standard techniques known in the art.

5 EXAMPLE 20: Expression of PRO Polypeptides in *E. coli*

This example illustrates preparation of an unglycosylated form of PRO by recombinant expression in *E. coli*.

The DNA sequence encoding PRO is initially amplified using selected PCR primers. The primers should contain restriction enzyme sites which correspond to the restriction enzyme sites on the selected 10 expression vector. A variety of expression vectors may be employed. An example of a suitable vector is pBR322 (derived from *E. coli*; see Bolivar et al., *Gene*, 2:95 (1977)) which contains genes for ampicillin and tetracycline resistance. The vector is digested with restriction enzyme and dephosphorylated. The PCR amplified sequences are then ligated into the vector. The vector will preferably include sequences which encode for an antibiotic resistance gene, a trp promoter, a polyhis leader (including the first six STII codons, polyhis sequence, and enterokinase cleavage site), the PRO coding region, lambda transcriptional terminator, and an argU gene.

The ligation mixture is then used to transform a selected *E. coli* strain using the methods described in Sambrook et al., *supra*. Transformants are identified by their ability to grow on LB plates and antibiotic resistant colonies are then selected. Plasmid DNA can be isolated and confirmed by restriction analysis and DNA sequencing.

Selected clones can be grown overnight in liquid culture medium such as LB broth supplemented with antibiotics. The overnight culture may subsequently be used to inoculate a larger scale culture. The cells are then grown to a desired optical density, during which the expression promoter is turned on.

20 After culturing the cells for several more hours, the cells can be harvested by centrifugation. The cell pellet obtained by the centrifugation can be solubilized using various agents known in the art, and the solubilized PRO protein can then be purified using a metal chelating column under conditions that allow tight binding of the protein.

PRO may be expressed in *E. coli* in a poly-His tagged form, using the following procedure. The DNA encoding PRO is initially amplified using selected PCR primers. The primers will contain restriction enzyme 30 sites which correspond to the restriction enzyme sites on the selected expression vector, and other useful sequences providing for efficient and reliable translation initiation, rapid purification on a metal chelation column, and proteolytic removal with enterokinase. The PCR-amplified, poly-His tagged sequences are then ligated into an expression vector, which is used to transform an *E. coli* host based on strain 52 (W3110 fuhA(tonA) lon galE rpoHts(htpRts) clpP(lacIq)). Transformants are first grown in LB containing 50 mg/ml 35 carbenicillin at 30°C with shaking until an O.D.600 of 3-5 is reached. Cultures are then diluted 50-100 fold into CRAP media (prepared by mixing 3.57 g (NH₄)₂SO₄, 0.71 g sodium citrate•2H₂O, 1.07 g KCl, 5.36 g Difco yeast extract, 5.36 g Sheffield hycase SF in 500 mL water, as well as 110 mM MPOS, pH 7.3, 0.55% (w/v)

glucose and 7 mM MgSO₄) and grown for approximately 20-30 hours at 30°C with shaking. Samples are removed to verify expression by SDS-PAGE analysis, and the bulk culture is centrifuged to pellet the cells. Cell pellets are frozen until purification and refolding.

5 *E. coli* paste from 0.5 to 1 L fermentations (6-10 g pellets) is resuspended in 10 volumes (w/v) in 7 M guanidine, 20 mM Tris, pH 8 buffer. Solid sodium sulfite and sodium tetrathionate is added to make final concentrations of 0.1M and 0.02 M, respectively, and the solution is stirred overnight at 4°C. This step results in a denatured protein with all cysteine residues blocked by sulfitolization. The solution is centrifuged at 40,000 rpm in a Beckman Ultracentrifuge for 30 min. The supernatant is diluted with 3-5 volumes of metal chelate column buffer (6 M guanidine, 20 mM Tris, pH 7.4) and filtered through 0.22 micron filters to clarify. The clarified extract is loaded onto a 5 ml Qiagen Ni-NTA metal chelate column equilibrated in the metal chelate column buffer. The column is washed with additional buffer containing 50 mM imidazole (Calbiochem, Utrol grade), pH 7.4. The protein is eluted with buffer containing 250 mM imidazole. Fractions containing the desired protein are pooled and stored at 4°C. Protein concentration is estimated by its absorbance at 280 nm using the calculated extinction coefficient based on its amino acid sequence.

15 The proteins are refolded by diluting the sample slowly into freshly prepared refolding buffer consisting of: 20 mM Tris, pH 8.6, 0.3 M NaCl, 2.5 M urea, 5 mM cysteine, 20 mM glycine and 1 mM EDTA. Refolding volumes are chosen so that the final protein concentration is between 50 to 100 micrograms/ml. The refolding solution is stirred gently at 4°C for 12-36 hours. The refolding reaction is quenched by the addition of TFA to a final concentration of 0.4% (pH of approximately 3). Before further purification of the protein, the solution is filtered through a 0.22 micron filter and acetonitrile is added to 2-10% final concentration. The refolded protein is chromatographed on a Poros R1/H reversed phase column using a mobile buffer of 0.1% TFA with elution with a gradient of acetonitrile from 10 to 80%. Aliquots of fractions with A280 absorbance are analyzed on SDS polyacrylamide gels and fractions containing homogeneous refolded protein are pooled. Generally, the properly refolded species of most proteins are eluted at the lowest concentrations of acetonitrile since those species are the most compact with their hydrophobic interiors shielded from interaction with the 20 reversed phase resin. Aggregated species are usually eluted at higher acetonitrile concentrations. In addition to resolving misfolded forms of proteins from the desired form, the reversed phase step also removes endotoxin 25 from the samples.

30 Fractions containing the desired folded PRO polypeptide are pooled and the acetonitrile removed using a gentle stream of nitrogen directed at the solution. Proteins are formulated into 20 mM Hepes, pH 6.8 with 0.14 M sodium chloride and 4% mannitol by dialysis or by gel filtration using G25 Superfine (Pharmacia) resins equilibrated in the formulation buffer and sterile filtered.

Many of the PRO polypeptides disclosed herein were successfully expressed as described above.

EXAMPLE 21: Expression of PRO Polypeptides in Mammalian Cells

35 This example illustrates preparation of a potentially glycosylated form of PRO by recombinant expression in mammalian cells.

The vector, pRK5 (see EP 307,247, published March 15, 1989), is employed as the expression vector.

Optionally, the PRO DNA is ligated into pRK5 with selected restriction enzymes to allow insertion of the PRO DNA using ligation methods such as described in Sambrook et al., supra. The resulting vector is called pRK5-PRO.

In one embodiment, the selected host cells may be 293 cells. Human 293 cells (ATCC CCL 1573) are 5 grown to confluence in tissue culture plates in medium such as DMEM supplemented with fetal calf serum and optionally, nutrient components and/or antibiotics. About 10 μ g pRK5-PRO DNA is mixed with about 1 μ g DNA encoding the VA RNA gene [Thimmappaya et al., Cell, 31:543 (1982)] and dissolved in 500 μ l of 1 mM Tris-HCl, 0.1 mM EDTA, 0.227 M CaCl₂. To this mixture is added, dropwise, 500 μ l of 50 mM HEPES (pH 7.35), 280 mM NaCl, 1.5 mM NaPO₄, and a precipitate is allowed to form for 10 minutes at 25°C. The 10 precipitate is suspended and added to the 293 cells and allowed to settle for about four hours at 37°C. The culture medium is aspirated off and 2 ml of 20% glycerol in PBS is added for 30 seconds. The 293 cells are then washed with serum free medium, fresh medium is added and the cells are incubated for about 5 days.

Approximately 24 hours after the transfections, the culture medium is removed and replaced with culture medium (alone) or culture medium containing 200 μ Ci/ml ³⁵S-cysteine and 200 μ Ci/ml ³⁵S-methionine. After 15 a 12 hour incubation, the conditioned medium is collected, concentrated on a spin filter, and loaded onto a 15% SDS gel. The processed gel may be dried and exposed to film for a selected period of time to reveal the presence of PRO polypeptide. The cultures containing transfected cells may undergo further incubation (in serum free medium) and the medium is tested in selected bioassays.

In an alternative technique, PRO may be introduced into 293 cells transiently using the dextran sulfate 20 method described by Somparyrac et al., Proc. Natl. Acad. Sci., 78:7575 (1981). 293 cells are grown to maximal density in a spinner flask and 700 μ g pRK5-PRO DNA is added. The cells are first concentrated from the spinner flask by centrifugation and washed with PBS. The DNA-dextran precipitate is incubated on the cell 25 pellet for four hours. The cells are treated with 20% glycerol for 90 seconds, washed with tissue culture medium, and re-introduced into the spinner flask containing tissue culture medium, 5 μ g/ml bovine insulin and 0.1 μ g/ml bovine transferrin. After about four days, the conditioned media is centrifuged and filtered to remove cells and debris. The sample containing expressed PRO can then be concentrated and purified by any selected method, such as dialysis and/or column chromatography.

In another embodiment, PRO can be expressed in CHO cells. The pRK5-PRO can be transfected into 30 CHO cells using known reagents such as CaPO₄ or DEAE-dextran. As described above, the cell cultures can be incubated, and the medium replaced with culture medium (alone) or medium containing a radiolabel such as ³⁵S-methionine. After determining the presence of PRO polypeptide, the culture medium may be replaced with serum free medium. Preferably, the cultures are incubated for about 6 days, and then the conditioned medium is harvested. The medium containing the expressed PRO can then be concentrated and purified by any selected 35 method.

Epitope-tagged PRO may also be expressed in host CHO cells. The PRO may be subcloned out of the pRK5 vector. The subclone insert can undergo PCR to fuse in frame with a selected epitope tag such as a poly-his tag into a Baculovirus expression vector. The poly-his tagged PRO insert can then be subcloned into a SV40

driven vector containing a selection marker such as DHFR for selection of stable clones. Finally, the CHO cells can be transfected (as described above) with the SV40 driven vector. Labeling may be performed, as described above, to verify expression. The culture medium containing the expressed poly-His tagged PRO can then be concentrated and purified by any selected method, such as by Ni^{2+} -chelate affinity chromatography.

PRO may also be expressed in CHO and/or COS cells by a transient expression procedure or in CHO 5 cells by another stable expression procedure.

Stable expression in CHO cells is performed using the following procedure. The proteins are expressed as an IgG construct (immunoadhesin), in which the coding sequences for the soluble forms (e.g. extracellular domains) of the respective proteins are fused to an IgG1 constant region sequence containing the hinge, CH2 and CH2 domains and/or is a poly-His tagged form.

10 Following PCR amplification, the respective DNAs are subcloned in a CHO expression vector using standard techniques as described in Ausubel et al., Current Protocols of Molecular Biology, Unit 3.16, John Wiley and Sons (1997). CHO expression vectors are constructed to have compatible restriction sites 5' and 3' of the DNA of interest to allow the convenient shuttling of cDNA's. The vector used expression in CHO cells is as described in Lucas et al., Nucl. Acids Res. 24:9 (1774-1779 (1996), and uses the SV40 early 15 promoter/enhancer to drive expression of the cDNA of interest and dihydrofolate reductase (DHFR). DHFR expression permits selection for stable maintenance of the plasmid following transfection.

12
16
20
24
28
32
36
40
44
48
52
56
60
64
68
72
76
80
84
88
92
96
100
104
108
112
116
120
124
128
132
136
140
144
148
152
156
160
164
168
172
176
180
184
188
192
196
200
204
208
212
216
220
224
228
232
236
240
244
248
252
256
260
264
268
272
276
280
284
288
292
296
300
304
308
312
316
320
324
328
332
336
340
344
348
352
356
360
364
368
372
376
380
384
388
392
396
400
404
408
412
416
420
424
428
432
436
440
444
448
452
456
460
464
468
472
476
480
484
488
492
496
500
504
508
512
516
520
524
528
532
536
540
544
548
552
556
560
564
568
572
576
580
584
588
592
596
600
604
608
612
616
620
624
628
632
636
640
644
648
652
656
660
664
668
672
676
680
684
688
692
696
700
704
708
712
716
720
724
728
732
736
740
744
748
752
756
760
764
768
772
776
780
784
788
792
796
800
804
808
812
816
820
824
828
832
836
840
844
848
852
856
860
864
868
872
876
880
884
888
892
896
900
904
908
912
916
920
924
928
932
936
940
944
948
952
956
960
964
968
972
976
980
984
988
992
996
1000
1004
1008
1012
1016
1020
1024
1028
1032
1036
1040
1044
1048
1052
1056
1060
1064
1068
1072
1076
1080
1084
1088
1092
1096
1100
1104
1108
1112
1116
1120
1124
1128
1132
1136
1140
1144
1148
1152
1156
1160
1164
1168
1172
1176
1180
1184
1188
1192
1196
1200
1204
1208
1212
1216
1220
1224
1228
1232
1236
1240
1244
1248
1252
1256
1260
1264
1268
1272
1276
1280
1284
1288
1292
1296
1300
1304
1308
1312
1316
1320
1324
1328
1332
1336
1340
1344
1348
1352
1356
1360
1364
1368
1372
1376
1380
1384
1388
1392
1396
1400
1404
1408
1412
1416
1420
1424
1428
1432
1436
1440
1444
1448
1452
1456
1460
1464
1468
1472
1476
1480
1484
1488
1492
1496
1500
1504
1508
1512
1516
1520
1524
1528
1532
1536
1540
1544
1548
1552
1556
1560
1564
1568
1572
1576
1580
1584
1588
1592
1596
1600
1604
1608
1612
1616
1620
1624
1628
1632
1636
1640
1644
1648
1652
1656
1660
1664
1668
1672
1676
1680
1684
1688
1692
1696
1700
1704
1708
1712
1716
1720
1724
1728
1732
1736
1740
1744
1748
1752
1756
1760
1764
1768
1772
1776
1780
1784
1788
1792
1796
1800
1804
1808
1812
1816
1820
1824
1828
1832
1836
1840
1844
1848
1852
1856
1860
1864
1868
1872
1876
1880
1884
1888
1892
1896
1900
1904
1908
1912
1916
1920
1924
1928
1932
1936
1940
1944
1948
1952
1956
1960
1964
1968
1972
1976
1980
1984
1988
1992
1996
2000
2004
2008
2012
2016
2020
2024
2028
2032
2036
2040
2044
2048
2052
2056
2060
2064
2068
2072
2076
2080
2084
2088
2092
2096
2100
2104
2108
2112
2116
2120
2124
2128
2132
2136
2140
2144
2148
2152
2156
2160
2164
2168
2172
2176
2180
2184
2188
2192
2196
2200
2204
2208
2212
2216
2220
2224
2228
2232
2236
2240
2244
2248
2252
2256
2260
2264
2268
2272
2276
2280
2284
2288
2292
2296
2300
2304
2308
2312
2316
2320
2324
2328
2332
2336
2340
2344
2348
2352
2356
2360
2364
2368
2372
2376
2380
2384
2388
2392
2396
2400
2404
2408
2412
2416
2420
2424
2428
2432
2436
2440
2444
2448
2452
2456
2460
2464
2468
2472
2476
2480
2484
2488
2492
2496
2500
2504
2508
2512
2516
2520
2524
2528
2532
2536
2540
2544
2548
2552
2556
2560
2564
2568
2572
2576
2580
2584
2588
2592
2596
2600
2604
2608
2612
2616
2620
2624
2628
2632
2636
2640
2644
2648
2652
2656
2660
2664
2668
2672
2676
2680
2684
2688
2692
2696
2700
2704
2708
2712
2716
2720
2724
2728
2732
2736
2740
2744
2748
2752
2756
2760
2764
2768
2772
2776
2780
2784
2788
2792
2796
2800
2804
2808
2812
2816
2820
2824
2828
2832
2836
2840
2844
2848
2852
2856
2860
2864
2868
2872
2876
2880
2884
2888
2892
2896
2900
2904
2908
2912
2916
2920
2924
2928
2932
2936
2940
2944
2948
2952
2956
2960
2964
2968
2972
2976
2980
2984
2988
2992
2996
3000
3004
3008
3012
3016
3020
3024
3028
3032
3036
3040
3044
3048
3052
3056
3060
3064
3068
3072
3076
3080
3084
3088
3092
3096
3100
3104
3108
3112
3116
3120
3124
3128
3132
3136
3140
3144
3148
3152
3156
3160
3164
3168
3172
3176
3180
3184
3188
3192
3196
3200
3204
3208
3212
3216
3220
3224
3228
3232
3236
3240
3244
3248
3252
3256
3260
3264
3268
3272
3276
3280
3284
3288
3292
3296
3300
3304
3308
3312
3316
3320
3324
3328
3332
3336
3340
3344
3348
3352
3356
3360
3364
3368
3372
3376
3380
3384
3388
3392
3396
3400
3404
3408
3412
3416
3420
3424
3428
3432
3436
3440
3444
3448
3452
3456
3460
3464
3468
3472
3476
3480
3484
3488
3492
3496
3500
3504
3508
3512
3516
3520
3524
3528
3532
3536
3540
3544
3548
3552
3556
3560
3564
3568
3572
3576
3580
3584
3588
3592
3596
3600
3604
3608
3612
3616
3620
3624
3628
3632
3636
3640
3644
3648
3652
3656
3660
3664
3668
3672
3676
3680
3684
3688
3692
3696
3700
3704
3708
3712
3716
3720
3724
3728
3732
3736
3740
3744
3748
3752
3756
3760
3764
3768
3772
3776
3780
3784
3788
3792
3796
3800
3804
3808
3812
3816
3820
3824
3828
3832
3836
3840
3844
3848
3852
3856
3860
3864
3868
3872
3876
3880
3884
3888
3892
3896
3900
3904
3908
3912
3916
3920
3924
3928
3932
3936
3940
3944
3948
3952
3956
3960
3964
3968
3972
3976
3980
3984
3988
3992
3996
4000
4004
4008
4012
4016
4020
4024
4028
4032
4036
4040
4044
4048
4052
4056
4060
4064
4068
4072
4076
4080
4084
4088
4092
4096
4100
4104
4108
4112
4116
4120
4124
4128
4132
4136
4140
4144
4148
4152
4156
4160
4164
4168
4172
4176
4180
4184
4188
4192
4196
4200
4204
4208
4212
4216
4220
4224
4228
4232
4236
4240
4244
4248
4252
4256
4260
4264
4268
4272
4276
4280
4284
4288
4292
4296
4300
4304
4308
4312
4316
4320
4324
4328
4332
4336
4340
4344
4348
4352
4356
4360
4364
4368
4372
4376
4380
4384
4388
4392
4396
4400
4404
4408
4412
4416
4420
4424
4428
4432
4436
4440
4444
4448
4452
4456
4460
4464
4468
4472
4476
4480
4484
4488
4492
4496
4500
4504
4508
4512
4516
4520
4524
4528
4532
4536
4540
4544
4548
4552
4556
4560
4564
4568
4572
4576
4580
4584
4588
4592
4596
4600
4604
4608
4612
4616
4620
4624
4628
4632
4636
4640
4644
4648
4652
4656
4660
4664
4668
4672
4676
4680
4684
4688
4692
4696
4700
4704
4708
4712
4716
4720
4724
4728
4732
4736
4740
4744
4748
4752
4756
4760
4764
4768
4772
4776
4780
4784
4788
4792
4796
4800
4804
4808
4812
4816
4820
4824
4828
4832
4836
4840
4844
4848
4852
4856
4860
4864
4868
4872
4876
4880
4884
4888
4892
4896
4900
4904
4908
4912
4916
4920
4924
4928
4932
4936
4940
4944
4948
4952
4956
4960
4964
4968
4972
4976
4980
4984
4988
4992
4996
5000
5004
5008
5012
5016
5020
5024
5028
5032
5036
5040
5044
5048
5052
5056
5060
5064
5068
5072
5076
5080
5084
5088
5092
5096
5100
5104
5108
5112
5116
5120
5124
5128
5132
5136
5140
5144
5148
5152
5156
5160
5164
5168
5172
5176
5180
5184
5188
5192
5196
5200
5204
5208
5212
5216
5220
5224
5228
5232
5236
5240
5244
5248
5252
5256
5260
5264
5268
5272
5276
5280
5284
5288
5292
5296
5300
5304
5308
5312
5316
5320
5324
5328
5332
5336
5340
5344
5348
5352
5356
5360
5364
5368
5372
5376
5380
5384
5388
5392
5396
5400
5404
5408
5412
5416
5420
5424
5428
5432
5436
5440
5444
5448
5452
5456
5460
5464
5468
5472
5476
5480
5484
5488
5492
5496
5500
5504
5508
5512
5516
5520
5524
5528
5532
5536
5540
5544
5548
5552
5556
5560
5564
5568
5572
5576
5580
5584
5588
5592
5596
5600
5604
5608
5612
5616
5620
5624
5628
5632
5636
5640
5644
5648
5652
5656
5660
5664
5668
5672
5676
5680
5684
5688
5692
5696
5700
5704
5708
5712
5716
5720
5724
5728
5732
5736
5740
5744
5748
5752
5756
5760
5764
5768
5772
5776
5780
5784
5788
5792
5796
5800
5804
5808
5812
5816
5820
5824
5828
5832
5836
5840
5844
5848
5852
5856
5860
5864
5868
5872
5876
5880
5884
5888
5892
5896
5900
5904
5908
5912
5916
5920
5924
5928
5932
5936
5940
5944
5948
5952
5956
5960
5964
5968
5972
5976
5980
5984
5988
5992
5996
6000
6004
6008
6012
6016
6020
6024
6028
6032
6036
6040
6044
6048
6052
6056
6060
6064
6068
6072
6076
6080
6084
6088
6092
6096
6100
6104
6108
6112
6116
6120
6124
6128
6132
6136
6140
6144
6148
6152
6156
6160
6164
6168
6172
6176
6180
6184
6188
6192
6196
6200
6204
6208
6212
6216
6220
6224
6228
6232
6236
6240
6244
6248
6252
6256
6260
6264
6268
6272
6276
6280
6284
6288
6292
6296
6300
6304
6308
6312
6316
6320
6324
6328
6332
6336
6340
6344
6348
6352
6356
6360
6364
6368
6372
6376
6380
6384
6388
6392
6396
6400
6404
6408
6412
6416
6420
6424
6428
6432
6436
6440
6444
6448
6452
6456
6460
6464
6468
6472
6476
6480
6484
6488
6492
6496
6500
6504
6508
6512
6516
6520
6524
6528
6532
6536
6540
6544
6548
6552
6556
6560
6564
6568
6572
6576
6580
6584
6588
6592
6596
6600
6604
6608
6612
6616
6620
6624
6628
6632
6636
6640
6644
6648
6652
6656
6660
6664
6668
6672
6676
6680
6684
6688
6692
6696
6700
6704
6708
6712
6716
6720
6724
6728
6732
6736
6740
6744
6748
6752
6756
6760
6764
6768
6772
6776
6780
6784
6788
6792
6796
6800
6804
6808
6812
6816
6820
6824
6828
6832
6836
6840
6844
6848
6852
6856
6860
6864
6868
6872
6876
6880
6884
6888
6892
6896
6900
6904
6908
6912
6916
6920
6924
6928
6932
6936
6940
6944
6948
6952
6956
6960
6964
6968
6972
6976
6980
6984
6988
6992
6996
7000
7004
7008
7012
7016
7020
7024
7028
7032
7036
7040
7044
7048
7052
7056
7060
7064
7068
7072
7076
7080
7084
7088
7092
7096
7100
7104
7108
7112
7116
7120
7124
7128
7132
7136
7140
7144
7148
7152
7156
7160
7164
7168
7172
7176
7180
7184
7188
7192
7196
7200
7204
7208
7212
7216
7220
7224
7228
7232
7236
7240
7244
7248
7252
7256
7260
7264
7268
7272
7276
7280
7284
7288
7292
7296
7300
7304
7308
7312
7316
7320
7324
7328
7332
7336
7340
7344
7348
7352
7356
7360
7364
7368
7372
7376
7380
7384
7388
7392
7396
7400
7404
7408
7412
7

For the poly-His tagged constructs, the proteins are purified using a Ni-NTA column (Qiagen). Before purification, imidazole is added to the conditioned media to a concentration of 5 mM. The conditioned media is pumped onto a 6 ml Ni-NTA column equilibrated in 20 mM Hepes, pH 7.4, buffer containing 0.3 M NaCl and 5 mM imidazole at a flow rate of 4-5 ml/min. at 4°C. After loading, the column is washed with additional equilibration buffer and the protein eluted with equilibration buffer containing 0.25 M imidazole. The highly purified protein is subsequently desalting into a storage buffer containing 10 mM Hepes, 0.14 M NaCl and 4% mannitol, pH 6.8, with a 25 ml G25 Superfine (Pharmacia) column and stored at -80°C.

Immunoadhesin (Fc-containing) constructs are purified from the conditioned media as follows. The conditioned medium is pumped onto a 5 ml Protein A column (Pharmacia) which had been equilibrated in 20 mM Na phosphate buffer, pH 6.8. After loading, the column is washed extensively with equilibration buffer before elution with 100 mM citric acid, pH 3.5. The eluted protein is immediately neutralized by collecting 1 ml fractions into tubes containing 275 μ L of 1 M Tris buffer, pH 9. The highly purified protein is subsequently desalting into storage buffer as described above for the poly-His tagged proteins. The homogeneity is assessed by SDS polyacrylamide gels and by N-terminal amino acid sequencing by Edman degradation.

Many of the PRO polypeptides disclosed herein were successfully expressed as described above.

EXAMPLE 22: Expression of PRO in Yeast

The following method describes recombinant expression of PRO in yeast.

First, yeast expression vectors are constructed for intracellular production or secretion of PRO from the ADH2/GAPDH promoter. DNA encoding PRO and the promoter is inserted into suitable restriction enzyme sites in the selected plasmid to direct intracellular expression of PRO. For secretion, DNA encoding PRO can be cloned into the selected plasmid, together with DNA encoding the ADH2/GAPDH promoter, a native PRO signal peptide or other mammalian signal peptide, or, for example, a yeast alpha-factor or invertase secretory signal/leader sequence, and linker sequences (if needed) for expression of PRO.

Yeast cells, such as yeast strain AB110, can then be transformed with the expression plasmids described above and cultured in selected fermentation media. The transformed yeast supernatants can be analyzed by precipitation with 10% trichloroacetic acid and separation by SDS-PAGE, followed by staining of the gels with Coomassie Blue stain.

Recombinant PRO can subsequently be isolated and purified by removing the yeast cells from the fermentation medium by centrifugation and then concentrating the medium using selected cartridge filters. The concentrate containing PRO may further be purified using selected column chromatography resins.

Many of the PRO polypeptides disclosed herein were successfully expressed as described above.

EXAMPLE 23: Expression of PRO in Baculovirus-Infected Insect Cells

The following method describes recombinant expression of PRO in Baculovirus-infected insect cells.

The sequence coding for PRO is fused upstream of an epitope tag contained within a baculovirus expression vector. Such epitope tags include poly-his tags and immunoglobulin tags (like Fc regions of IgG). A variety of plasmids may be employed, including plasmids derived from commercially available plasmids such

as pVL1393 (Novagen). Briefly, the sequence encoding PRO or the desired portion of the coding sequence of PRO such as the sequence encoding the extracellular domain of a transmembrane protein or the sequence encoding the mature protein if the protein is extracellular is amplified by PCR with primers complementary to the 5' and 3' regions. The 5' primer may incorporate flanking (selected) restriction enzyme sites. The product is then digested with those selected restriction enzymes and subcloned into the expression vector.

5 Recombinant baculovirus is generated by co-transfected the above plasmid and BaculoGoldTM virus DNA (Pharmingen) into *Spodoptera frugiperda* ("Sf9") cells (ATCC CRL 1711) using lipofectin (commercially available from GIBCO-BRL). After 4 - 5 days of incubation at 28°C, the released viruses are harvested and used for further amplifications. Viral infection and protein expression are performed as described by O'Reilley et al., Baculovirus expression vectors: A Laboratory Manual, Oxford: Oxford University Press (1994).

10 Expressed poly-his tagged PRO can then be purified, for example, by Ni²⁺-chelate affinity chromatography as follows. Extracts are prepared from recombinant virus-infected Sf9 cells as described by Rupert et al., Nature, 362:175-179 (1993). Briefly, Sf9 cells are washed, resuspended in sonication buffer (25 mL Hepes, pH 7.9; 12.5 mM MgCl₂; 0.1 mM EDTA; 10% glycerol; 0.1% NP-40; 0.4 M KCl), and sonicated twice for 20 seconds on ice. The sonicates are cleared by centrifugation, and the supernatant is diluted 50-fold in loading buffer (50 mM phosphate, 300 mM NaCl, 10% glycerol, pH 7.8) and filtered through a 0.45 μ m filter. A Ni²⁺-NTA agarose column (commercially available from Qiagen) is prepared with a bed volume of 5 mL, washed with 25 mL of water and equilibrated with 25 mL of loading buffer. The filtered cell extract is loaded onto the column at 0.5 mL per minute. The column is washed to baseline A₂₈₀ with loading buffer, at which point fraction collection is started. Next, the column is washed with a secondary wash buffer (50 mM phosphate; 300 mM NaCl, 10% glycerol, pH 6.0), which elutes nonspecifically bound protein. After reaching A₂₈₀ baseline again, the column is developed with a 0 to 500 mM Imidazole gradient in the secondary wash buffer. One mL fractions are collected and analyzed by SDS-PAGE and silver staining or Western blot with Ni²⁺-NTA-conjugated to alkaline phosphatase (Qiagen). Fractions containing the eluted His₁₀-tagged PRO are pooled and dialyzed against loading buffer.

25 Alternatively, purification of the IgG tagged (or Fc tagged) PRO can be performed using known chromatography techniques, including for instance, Protein A or protein G column chromatography.

Many of the PRO polypeptides disclosed herein were successfully expressed as described above.

EXAMPLE 24: Preparation of Antibodies that Bind PRO

30 This example illustrates preparation of monoclonal antibodies which can specifically bind PRO.

Techniques for producing the monoclonal antibodies are known in the art and are described, for instance, in Goding, supra. Immunogens that may be employed include purified PRO, fusion proteins containing PRO, and cells expressing recombinant PRO on the cell surface. Selection of the immunogen can be made by the skilled artisan without undue experimentation.

35 Mice, such as Balb/c, are immunized with the PRO immunogen emulsified in complete Freund's adjuvant and injected subcutaneously or intraperitoneally in an amount from 1-100 micrograms. Alternatively, the immunogen is emulsified in MPL-TDM adjuvant (Ribi Immunochemical Research, Hamilton, MT) and

injected into the animal's hind foot pads. The immunized mice are then boosted 10 to 12 days later with additional immunogen emulsified in the selected adjuvant. Thereafter, for several weeks, the mice may also be boosted with additional immunization injections. Serum samples may be periodically obtained from the mice by retro-orbital bleeding for testing in ELISA assays to detect anti-PRO antibodies.

After a suitable antibody titer has been detected, the animals "positive" for antibodies can be injected

5 with a final intravenous injection of PRO. Three to four days later, the mice are sacrificed and the spleen cells are harvested. The spleen cells are then fused (using 35% polyethylene glycol) to a selected murine myeloma cell line such as P3X63AgU.1, available from ATCC, No. CRL 1597. The fusions generate hybridoma cells which can then be plated in 96 well tissue culture plates containing HAT (hypoxanthine, aminopterin, and thymidine) medium to inhibit proliferation of non-fused cells, myeloma hybrids, and spleen cell hybrids.

10 The hybridoma cells will be screened in an ELISA for reactivity against PRO. Determination of "positive" hybridoma cells secreting the desired monoclonal antibodies against PRO is within the skill in the art.

15 The positive hybridoma cells can be injected intraperitoneally into syngeneic Balb/c mice to produce ascites containing the anti-PRO monoclonal antibodies. Alternatively, the hybridoma cells can be grown in tissue culture flasks or roller bottles. Purification of the monoclonal antibodies produced in the ascites can be accomplished using ammonium sulfate precipitation, followed by gel exclusion chromatography. Alternatively, affinity chromatography based upon binding of antibody to protein A or protein G can be employed.

EXAMPLE 25: Purification of PRO Polypeptides Using Specific Antibodies

Native or recombinant PRO polypeptides may be purified by a variety of standard techniques in the art 20 of protein purification. For example, pro-PRO polypeptide, mature PRO polypeptide, or pre-PRO polypeptide is purified by immunoaffinity chromatography using antibodies specific for the PRO polypeptide of interest. In general, an immunoaffinity column is constructed by covalently coupling the anti-PRO polypeptide antibody to an activated chromatographic resin.

25 Polyclonal immunoglobulins are prepared from immune sera either by precipitation with ammonium sulfate or by purification on immobilized Protein A (Pharmacia LKB Biotechnology, Piscataway, N.J.). Likewise, monoclonal antibodies are prepared from mouse ascites fluid by ammonium sulfate precipitation or chromatography on immobilized Protein A. Partially purified immunoglobulin is covalently attached to a chromatographic resin such as CnBr-activated SEPHAROSE™ (Pharmacia LKB Biotechnology). The antibody is coupled to the resin, the resin is blocked, and the derivative resin is washed according to the manufacturer's 30 instructions.

Such an immunoaffinity column is utilized in the purification of PRO polypeptide by preparing a fraction from cells containing PRO polypeptide in a soluble form. This preparation is derived by solubilization of the whole cell or of a subcellular fraction obtained via differential centrifugation by the addition of detergent or by other methods well known in the art. Alternatively, soluble PRO polypeptide containing a signal sequence may 35 be secreted in useful quantity into the medium in which the cells are grown.

A soluble PRO polypeptide-containing preparation is passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PRO polypeptide (*e.g.*, high ionic

strength buffers in the presence of detergent). Then, the column is eluted under conditions that disrupt antibody/PRO polypeptide binding (*e.g.*, a low pH buffer such as approximately pH 2-3, or a high concentration of a chaotrope such as urea or thiocyanate ion), and PRO polypeptide is collected.

EXAMPLE 26: Drug Screening

5 This invention is particularly useful for screening compounds by using PRO polypeptides or binding fragment thereof in any of a variety of drug screening techniques. The PRO polypeptide or fragment employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the PRO polypeptide or fragment. Drugs are screened
10 against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between PRO polypeptide or a fragment and the agent being tested. Alternatively, one can examine the diminution in complex formation between the PRO polypeptide and its target cell or target receptors caused by the agent being tested.

15 Thus, the present invention provides methods of screening for drugs or any other agents which can affect a PRO polypeptide-associated disease or disorder. These methods comprise contacting such an agent with an PRO polypeptide or fragment thereof and assaying (i) for the presence of a complex between the agent and the PRO polypeptide or fragment, or (ii) for the presence of a complex between the PRO polypeptide or fragment and the cell, by methods well known in the art. In such competitive binding assays, the PRO polypeptide or fragment is typically labeled. After suitable incubation, free PRO polypeptide or fragment is separated from that present in bound form, and the amount of free or uncomplexed label is a measure of the ability of the particular agent to bind to PRO polypeptide or to interfere with the PRO polypeptide/cell complex.

20 Another technique for drug screening provides high throughput screening for compounds having suitable binding affinity to a polypeptide and is described in detail in WO 84/03564, published on September 13, 1984. Briefly stated, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. As applied to a PRO polypeptide, the peptide test compounds are reacted with PRO polypeptide and washed. Bound PRO polypeptide is detected by methods well known in the art. Purified PRO polypeptide can also be coated directly onto plates for use in the aforementioned drug screening techniques. In addition, non-neutralizing antibodies can be used to capture the peptide and immobilize it on the solid support.

25 30 This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of binding PRO polypeptide specifically compete with a test compound for binding to PRO polypeptide or fragments thereof. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PRO polypeptide.

EXAMPLE 27: Rational Drug Design

The goal of rational drug design is to produce structural analogs of biologically active polypeptide of interest (*i.e.*, a PRO polypeptide) or of small molecules with which they interact, *e.g.*, agonists, antagonists, or

inhibitors. Any of these examples can be used to fashion drugs which are more active or stable forms of the PRO polypeptide or which enhance or interfere with the function of the PRO polypeptide *in vivo* (c.f., Hodgson, *Bio/Technology*, 9: 19-21 (1991)).

In one approach, the three-dimensional structure of the PRO polypeptide, or of an PRO polypeptide-inhibitor complex, is determined by x-ray crystallography, by computer modeling or, most typically, 5 by a combination of the two approaches. Both the shape and charges of the PRO polypeptide must be ascertained to elucidate the structure and to determine active site(s) of the molecule. Less often, useful information regarding the structure of the PRO polypeptide may be gained by modeling based on the structure of homologous proteins. In both cases, relevant structural information is used to design analogous PRO polypeptide-like molecules or to identify efficient inhibitors. Useful examples of rational drug design may include molecules which have improved 10 activity or stability as shown by Braxton and Wells, *Biochemistry*, 31:7796-7801 (1992) or which act as inhibitors, agonists, or antagonists of native peptides as shown by Athauda *et al.*, *J. Biochem.*, 113:742-746 (1993).

It is also possible to isolate a target-specific antibody, selected by functional assay, as described above, and then to solve its crystal structure. This approach, in principle, yields a pharmacore upon which subsequent 15 drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idiotypic antibodies (anti-ids) to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an analog of the original receptor. The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced peptides. The isolated peptides would then act as the pharmacore.

20 By virtue of the present invention, sufficient amounts of the PRO polypeptide may be made available to perform such analytical studies as X-ray crystallography. In addition, knowledge of the PRO polypeptide amino acid sequence provided herein will provide guidance to those employing computer modeling techniques in place of or in addition to x-ray crystallography.

25 **EXAMPLE 28: Gene Amplification**

This example shows that the PRO327-, PRO344-, PRO347-, PRO357- and PRO715-encoding genes are amplified in the genome of certain human lung, colon and/or breast cancers and/or cell lines. Amplification is associated with overexpression of the gene product, indicating that the polypeptides are useful targets for therapeutic intervention in certain cancers such as colon, lung, breast and other cancers. Therapeutic agents may 30 take the form of antagonists of PRO327, PRO344, PRO347, PRO357 or PRO715 polypeptide, for example, murine-human chimeric, humanized or human antibodies against a PRO327, PRO344, PRO347, PRO357 or PRO715 polypeptide. These amplifications also are useful as diagnostic markers for the presence of a specific type of tumor type.

The starting material for the screen was genomic DNA isolated from a variety cancers. The DNA is 35 quantitated precisely, *e.g.*, fluorometrically. As a negative control, DNA was isolated from the cells of ten normal healthy individuals which was pooled and used as assay controls for the gene copy in healthy individuals (not shown). The 5' nuclease assay (for example, TaqManTM) and real-time quantitative PCR (for example, ABI

Prizm 7700 Sequence Detection System™ (Perkin Elmer, Applied Biosystems Division, Foster City, CA)), were used to find genes potentially amplified in certain cancers. The results were used to determine whether the DNA encoding PRO327, PRO344, PRO347, PRO357 or PRO715 is over-represented in any of the primary lung or colon cancers or cancer cell lines or breast cancer cell lines that were screened. The primary lung cancers were obtained from individuals with tumors of the type and stage as indicated in Table 9. An explanation of the 5 abbreviations used for the designation of the primary tumors listed in Table 9 and the primary tumors and cell lines referred to throughout this example are given below.

The results of the TaqMan™ are reported in delta (Δ) Ct units. One unit corresponds to 1 PCR cycle or approximately a 2-fold amplification relative to normal, two units corresponds to 4-fold, 3 units to 8-fold amplification and so on. Quantitation was obtained using primers and a TaqMan™ fluorescent probe derived 10 from the PRO327-, PRO344-, PRO347-, PRO357- or PRO715-encoding gene. Regions of PRO327, PRO344, PRO347, PRO357 or PRO715 which are most likely to contain unique nucleic acid sequences and which are least likely to have spliced out introns are preferred for the primer and probe derivation, *e.g.*, 3'-untranslated regions. The sequences for the primers and probes (forward, reverse and probe) used for the PRO327, PRO344, PRO347, PRO357 or PRO715 gene amplification analysis were as follows:

PRO327 (DNA38113-1230)

<u>forward</u>	5'-CTCAAGAACGACGCGTACTGC-3'	(SEQ ID NO:96)
<u>probe</u>	5'-CCAACCTCAGCTTCCGCCTACGA-3'	(SEQ ID NO:97)
<u>reverse</u>	5'-CATCCAGGCTGCCACTG-3'	(SEQ ID NO:98)

PRO344 (DNA40592-1242)

<u>forward</u>	5'-TGGCAAGGAATGGGAACAGT-3'	(SEQ ID NO:99)
<u>probe</u>	5'-ATGCTGCCAGACCTGATCGCAGACA-3'	(SEQ ID NO:100)
<u>reverse</u>	5'-GGGCAGAAATCCAGCCACT-3'	(SEQ ID NO:101)

PRO347 (DNA44176-1244)

<u>forward</u>	5'-CCCTCGCCTGCTTTGA-3'	(SEQ ID NO:102)
<u>probe</u>	5'-GCCATCTAATTGAAGCCCATCTTCCA-3'	(SEQ ID NO:103)
<u>reverse</u>	5'-CTGGCGGTGTCCTCTCCTT-3'	(SEQ ID NO:104)

PRO357 (DNA44804-1248)

<u>forward</u>	5'-CCTCGGTCTCCTCATCTGTGA-3'	(SEQ ID NO:105)
<u>probe</u>	5'-TGGCCCAGCTGACGAGCCCT-3'	(SEQ ID NO:106)
<u>reverse</u>	5'-CTCATAGGCACTCGGTTCTGG-3'	(SEQ ID NO:107)

PRO715 (DNA52722-1229)

<u>forward</u>	5'-TGGCTCCCAGCTTGGAAAGA-3'	(SEQ ID NO:108)
<u>probe</u>	5'-CAGCTCTTGGCTGTCTCCAGTATGTACCCA-3'	(SEQ ID NO:109)
<u>reverse</u>	5'-GATGCCTCTGTTCCCTGCACAT-3'	(SEQ ID NO:110)

The 5' nuclease assay reaction is a fluorescent PCR-based technique which makes use of the 5' exonuclease activity of Taq DNA polymerase enzyme to monitor amplification in real time. Two oligonucleotide primers are used to generate an amplicon typical of a PCR reaction. A third oligonucleotide, or probe, is designed to detect nucleotide sequence located between the two PCR primers. The probe is non-extendible by Taq DNA polymerase enzyme, and is labeled with a reporter fluorescent dye and a quencher fluorescent dye. Any laser-induced emission from the reporter dye is quenched by the quenching dye when the two dyes are located close together as they are on the probe. During the amplification reaction, the Taq DNA polymerase enzyme cleaves the probe in a template-dependent manner. The resultant probe fragments disassociate in solution, and signal from the released reporter dye is free from the quenching effect of the second fluorophore. One molecule of reporter dye is liberated for each new molecule synthesized, and detection of the unquenched reporter dye provides the basis for quantitative interpretation of the data.

15 The 5' nuclease procedure is run on a real-time quantitative PCR device such as the ABI Prism 7700TM Sequence Detection. The system consists of a thermocycler, laser, charge-coupled device (CCD) camera and computer. The system amplifies samples in a 96-well format on a thermocycler. During amplification, laser-induced fluorescent signal is collected in real-time through fiber optics cables for all 96 wells, and detected at the CCD. The system includes software for running the instrument and for analyzing the data.

20 5' Nuclease assay data are initially expressed as Ct, or the threshold cycle. This is defined as the cycle at which the reporter signal accumulates above the background level of fluorescence. The ΔCt values are used as quantitative measurement of the relative number of starting copies of a particular target sequence in a nucleic acid sample when comparing cancer DNA results to normal human DNA results.

25 Table 9 describes the stage, T stage and N stage of various primary tumors which were used to screen the PRO327, PRO344, PRO347, PRO357 and PRO715 compounds of the invention.

Table 9
Primary Lung and Colon Tumor Profiles

	<u>Primary Tumor Stage</u>	<u>Stage</u>	<u>Other Stage</u>	<u>Dukes Stage</u>	<u>T Stage</u>	<u>N Stage</u>
5	Human lung tumor AdenoCa (SRCC724) [LT1]	IIA			T1	N1
	Human lung tumor SqCCa (SRCC725) [LT1a]	IIB			T3	N0
	Human lung tumor AdenoCa (SRCC726) [LT2]	IB			T2	N0
	Human lung tumor AdenoCa (SRCC727) [LT3]	IIIA			T1	N2
	Human lung tumor AdenoCa (SRCC728) [LT4]	IB			T2	N0
	Human lung tumor SqCCa (SRCC729) [LT6]	IB			T2	N0
10	Human lung tumor Aden/SqCCa (SRCC730) [LT7]	IA			T1	N0
	Human lung tumor AdenoCa (SRCC731) [LT9]	IB			T2	N0
	Human lung tumor SqCCa (SRCC732) [LT10]	IIB			T2	N1
	Human lung tumor SqCCa (SRCC733) [LT11]	IIA			T1	N1
	Human lung tumor AdenoCa (SRCC734) [LT12]	IV			T2	N0
15	Human lung tumor AdenoSqCCa (SRCC735) [LT13]	IB			T2	N0
	Human lung tumor SqCCa (SRCC736) [LT15]	IB			T2	N0
	Human lung tumor SqCCa (SRCC737) [LT16]	IB			T2	N0
	Human lung tumor SqCCa (SRCC738) [LT17]	IIB			T2	N1
	Human lung tumor SqCCa (SRCC739) [LT18]	IB			T2	N0
20	Human lung tumor SqCCa (SRCC740) [LT19]	IB			T2	N0
	Human lung tumor LCCa (SRCC741) [LT21]	IIB			T3	N1
	Human lung AdenoCa (SRCC811) [LT22]	IA			T1	N0
	Human colon AdenoCa (SRCC742) [CT2]		M1	D	pT4	N0
	Human colon AdenoCa (SRCC743) [CT3]			B	pT3	N0
25	Human colon AdenoCa (SRCC744) [CT8]			B	T3	N0
	Human colon AdenoCa (SRCC745) [CT10]			A	pT2	N0
	Human colon AdenoCa (SRCC746) [CT12]		MO, R1	B	T3	N0
	Human colon AdenoCa (SRCC747) [CT14]		pMO, RO	B	pT3	pN0
	Human colon AdenoCa (SRCC748) [CT15]		M1, R2	D	T4	N2
30	Human colon AdenoCa (SRCC749) [CT16]		pMO	B	pT3	pN0
	Human colon AdenoCa (SRCC750) [CT17]			C1	pT3	pN1
	Human colon AdenoCa (SRCC751) [CT1]		MO, R1	B	pT3	N0
	Human colon AdenoCa (SRCC752) [CT4]			B	pT3	M0
	Human colon AdenoCa (SRCC753) [CT5]		G2	C1	pT3	pN0
35	Human colon AdenoCa (SRCC754) [CT6]		pMO, RO	B	pT3	pN0
	Human colon AdenoCa (SRCC755) [CT7]		G1	A	pT2	pN0
	Human colon AdenoCa (SRCC756) [CT9]		G3	D	pT4	pN2
	Human colon AdenoCa (SRCC757) [CT11]			B	T3	N0
	Human colon AdenoCa (SRCC758) [CT18]		MO, RO	B	pT3	pN0
40						

DNA Preparation:

DNA was prepared from cultured cell lines, primary tumors, normal human blood. The isolation was performed using purification kit, buffer set and protease and all from Quiagen, according to the manufacturer's instructions and the description below.

Cell culture lysis:

Cells were washed and trypsinized at a concentration of 7.5×10^8 per tip and pelleted by centrifuging at 1000 rpm for 5 minutes at 4°C, followed by washing again with 1/2 volume of PBS re-centrifugation. The pellets were washed a third time, the suspended cells collected and washed 2x with PBS. The cells were then suspended into 10 ml PBS. Buffer C1 was equilibrated at 4°C. Qiagen protease #19155 was diluted into 6.25 ml cold ddH₂O to a final concentration of 20 mg/ml and equilibrated at 4°C. 10 ml of G2 Buffer was prepared

by diluting Qiagen RNase A stock (100 mg/ml) to a final concentration of 200 μ g/ml.

Buffer C1 (10 ml, 4°C) and ddH₂O (40 ml, 4°C) were then added to the 10 ml of cell suspension, mixed by inverting and incubated on ice for 10 minutes. The cell nuclei were pelleted by centrifuging in a Beckman swinging bucket rotor at 2500 rpm at 4°C for 15 minutes. The supernatant was discarded and the nuclei were suspended with a vortex into 2 ml Buffer C1 (at 4°C) and 6 ml ddH₂O, followed by a second 4°C centrifugation 5 at 2500 rpm for 15 minutes. The nuclei were then resuspended into the residual buffer using 200 μ l per tip. G2 buffer (10 ml) was added to the suspended nuclei while gentle vortexing was applied. Upon completion of buffer addition, vigorous vortexing was applied for 30 seconds. Quiagen protease (200 μ l, prepared as indicated above) was added and incubated at 50°C for 60 minutes. The incubation and centrifugation was repeated until the lysates were clear (e.g., incubating additional 30-60 minutes, pelleting at 3000 x g for 10 min., 4°C).

10 *Solid human tumor sample preparation and lysis:*

Tumor samples were weighed and placed into 50 ml conical tubes and held on ice. Processing was limited to no more than 250 mg tissue per preparation (1 tip/preparation). The protease solution was freshly prepared by diluting into 6.25 ml cold ddH₂O to a final concentration of 20 mg/ml and stored at 4°C. G2 buffer (20 ml) was prepared by diluting DNase A to a final concentration of 200 mg/ml (from 100 mg/ml stock). The tumor tissue was homogenated in 19 ml G2 buffer for 60 seconds using the large tip of the polytron in a laminar-flow TC hood in order to avoid inhalation of aerosols, and held at room temperature. Between samples, the polytron was cleaned by spinning at 2 x 30 seconds each in 2L ddH₂O, followed by G2 buffer (50 ml). If tissue was still present on the generator tip, the apparatus was disassembled and cleaned.

Quiagen protease (prepared as indicated above, 1.0 ml) was added, followed by vortexing and incubation at 50°C for 3 hours. The incubation and centrifugation was repeated until the lysates were clear (e.g., incubating additional 30-60 minutes, pelleting at 3000 x g for 10 min., 4°C).

15 *Human blood preparation and lysis:*

Blood was drawn from healthy volunteers using standard infectious agent protocols and citrated into 10 ml samples per tip. Quiagen protease was freshly prepared by dilution into 6.25 ml cold ddH₂O to a final 20 concentration of 20 mg/ml and stored at 4°C. G2 buffer was prepared by diluting RNase A to a final concentration of 200 μ g/ml from 100 mg/ml stock. The blood (10 ml) was placed into a 50 ml conical tube and 25 10 ml C1 buffer and 30 ml ddH₂O (both previously equilibrated to 4°C) were added, and the components mixed by inverting and held on ice for 10 minutes. The nuclei were pelleted with a Beckman swinging bucket rotor at 2500 rpm, 4°C for 15 minutes and the supernatant discarded. With a vortex, the nuclei were suspended into 30 2 ml C1 buffer (4°C) and 6 ml ddH₂O (4°C). Vortexing was repeated until the pellet was white. The nuclei were then suspended into the residual buffer using a 200 μ l tip. G2 buffer (10 ml) were added to the suspended nuclei while gently vortexing, followed by vigorous vortexing for 30 seconds. Quiagen protease was added (200 μ l) and incubated at 50°C for 60 minutes. The incubation and centrifugation was repeated until the lysates were clear (e.g., incubating additional 30-60 minutes, pelleting at 3000 x g for 10 min., 4°C).

35 *Purification of cleared lysates:*

(1) *Isolation of genomic DNA:*

Genomic DNA was equilibrated (1 sample per maxi tip preparation) with 10 ml QBT buffer. QF elution

buffer was equilibrated at 50°C. The samples were vortexed for 30 seconds, then loaded onto equilibrated tips and drained by gravity. The tips were washed with 2 x 15 ml QC buffer. The DNA was eluted into 30 ml silanized, autoclaved 30 ml Corex tubes with 15 ml QF buffer (50°C). Isopropanol (10.5 ml) was added to each sample, the tubes covered with parafin and mixed by repeated inversion until the DNA precipitated. Samples were pelleted by centrifugation in the SS-34 rotor at 15,000 rpm for 10 minutes at 4°C. The pellet location was marked, the supernatant discarded, and 10 ml 70% ethanol (4°C) was added. Samples were pelleted again by centrifugation on the SS-34 rotor at 10,000 rpm for 10 minutes at 4°C. The pellet location was marked and the supernatant discarded. The tubes were then placed on their side in a drying rack and dried 10 minutes at 37°C, taking care not to overdry the samples.

After drying, the pellets were dissolved into 1.0 ml TE (pH 8.5) and placed at 50°C for 1-2 hours. Samples were held overnight at 4°C as dissolution continued. The DNA solution was then transferred to 1.5 ml tubes with a 26 gauge needle on a tuberculin syringe. The transfer was repeated 5x in order to shear the DNA. Samples were then placed at 50°C for 1-2 hours.

(2) Quantitation of genomic DNA and preparation for gene amplification assay:

The DNA levels in each tube were quantified by standard A_{260} , A_{280} spectrophotometry on a 1:20 dilution (5 μ l DNA + 95 μ l ddH₂O) using the 0.1 ml quartz cuvetts in the Beckman DU640 spectrophotometer. A_{260}/A_{280} ratios were in the range of 1.8-1.9. Each DNA samples was then diluted further to approximately 200 ng/ml in TE (pH 8.5). If the original material was highly concentrated (about 700 ng/ μ l), the material was placed at 50°C for several hours until resuspended.

Fluorometric DNA quantitation was then performed on the diluted material (20-600 ng/ml) using the manufacturer's guidelines as modified below. This was accomplished by allowing a Hoeffer DyNA Quant 200 fluorometer to warm-up for about 15 minutes. The Hoechst dye working solution (#H33258, 10 μ l, prepared within 12 hours of use) was diluted into 100 ml 1 x TNE buffer. A 2 ml cuvette was filled with the fluorometer solution, placed into the machine, and the machine was zeroed. pGEM 3Zf(+) (2 μ l, lot #360851026) was added to 2 ml of fluorometer solution and calibrated at 200 units. An additional 2 μ l of pGEM 3Zf(+) DNA was then tested and the reading confirmed at 400 +/- 10 units. Each sample was then read at least in triplicate. When 3 samples were found to be within 10% of each other, their average was taken and this value was used as the quantification value.

The fluorometricly determined concentration was then used to dilute each sample to 10 ng/ μ l in ddH₂O. This was done simultaneously on all template samples for a single TaqMan plate assay, and with enough material to run 500-1000 assays. The samples were tested in triplicate with Taqman™ primers and probe both B-actin and GAPDH on a single plate with normal human DNA and no-template controls. The diluted samples were used provided that the CT value of normal human DNA subtracted from test DNA was +/- 1 Ct. The diluted, lot-qualified genomic DNA was stored in 1.0 ml aliquots at -80°C. Aliquots which were subsequently to be used in the gene amplification assay were stored at 4°C. Each 1 ml aliquot is enough for 8-9 plates or 64 tests.

35 *Gene amplification assay:*

The PRO327, PRO344, PRO347, PRO357 and PRO715 compounds of the invention were screened in the following primary tumors and the resulting Δ Ct values greater than or equal to 1.0 are reported in Table 10.

Table 10
 ΔCt values in primary tumor and cell lines models

Primary Tumors or Cell Lines		PRO327	PRO344	PRO347	PRO357	PRO715
5	LT1	---	---	1.035	---	1.625
LT1a		1.045	---	1.865 1.0 1.93	1.18 2.47	1.045
10	LT3	1.135	---	1.325 1.2	2.93	---
LT6		1.395	---	1.945 1.42	2.6 3.18	---
15	LT9	---	---	2.645	3.47 2.91	1.005
20	LT10	1.305	---	1.845 1.13	3.42 3.51	1.125
LT11		1.53	1.52	1.395 1.35	1.185 2.875	1.75 1.12
25	LT12	2.99 2.15	1.2	1.425 1.73	1.225 2.225	1.63 1.11 1.14
30	LT13	2.48 1.69	1.81 1.175 1.05	2.035 2.28 1.15	1.585 1.665 1.31	2.29 1.83
35	LT15	3.99 2.89	1.62 1.33	1.615 2.73 1.27 1.44	2.205 2.445 1.89	2.33 1.89
40	LT16	1.16 2.65	1.13 1.09	---	2.605	1.2 1.1
LT17		1.76 1.09	1.46	1.24	1.275 2.855 1.01	1.95 1.33
45	LT18	---	---	---	2.455	1.14
LT19		3.58	2.47	1.835 1.35	2.295 2.645	2.38
50	LT21	---	1.09	1.14	2.675	---
CT2		3.645	1.84	2.1 1.605	2.01	1.675 1.605
55	CT3	1.125	---	1.01	---	1.135 1.105

Table 10 (cont⁷)
ΔCt values in primary tumor and cell lines models

Primary Tumors or Cell Lines		PRO327	PRO344	PRO347	PRO357	PRO715
5	CT8	1.645	---	1.3	1.1	1.285
						1.345
10	CT10	2.535	---	---	1.42	2.155
						1.785
15	CT12	1.885	---	---	---	---
	CT14	2.515	1.16	1.39	1.5	1.265
				1.45		1.575
20	CT15	1.305	1.17	1.3	1.25	1.585
						1.475
25	CT16	1.475	---	1.33	1.05	1.095
				1.055		1.475
30	CT17	1.715	---	---	---	1.245
						1.375
35	CT1	1.375	1.245	1.045	1.045	1.285
						1.6
						1.085
	CT4	2.225	1.465	---	1.275	1.375
						2.23
						1.165
40	CT5	2.505	1.515	1.625	1.695	1.975
				1.985		2.07
						1.715
45	CT6	2.285	---	---	1.085	1.305
						1.73
						1.245

Table 10 (cont')
 ΔCt values in primary tumor and cell lines models

Primary Tumors or Cell Lines		PRO327	PRO344	PRO347	PRO357	PRO715
	CT7	---	---	---	1.735	1.005
5						1.65
						1.025
	CT9	1.585	---	---	---	1.0
10	CT11	3.335	1.355	1.315	1.835	2.185
				1.525		2.54
	CT18	1.075	---	---	---	1.69
15	SRCC771 (H157)	1.65	---	---	---	---
	SRCC772 (H441)	2.23	---	---	---	---
20	SRCC773 (H460)	1.12	---	---	---	---
	SRCC774 (SKMES-1)	1.18	---	---	---	---
	SRCC777 (SW620)	2.24	---	---	---	---
25	SRCC778 (Colo320)	1.01	---	---	---	---
	SRCC830 (HCC2998)	1.23	---	---	---	---
	SRCC831 (KM12)	1.61	---	---	---	---
30	SRCC832 (H522)	1.02	---	---	---	---
	SRCC833 (H810)	1.11	---	---	---	---

PRO327:

PRO327 (DNA38113-1230) was reexamined along with selected tumors from the above initial screen with framework mapping. Table 11 describes the framework markers that were employed in association with PRO327 (DNA38113-1230). The framework markers are located approximately every 20 megabases along Chromosome 19, and are used to control aneuploidy. The ΔCt values for the described framework markers along Chromosome 19 relative to PRO327 (DNA38113-1230) are indicated for selected tumors in Table 13.

PRO327 (DNA38113-1230) was also reexamined along with selected tumors from the above initial screen with epicenter mapping. Table 12 describes the epicenter markers that were employed in association with PRO327 (DNA38113-1230). These markers are located in close proximity to DNA38113-1230 and are used to assess the amplification status of the region of Chromosome 19 in which DNA38113-1230 is located. The distance between markers is measured in centirays (cR), which is a radiation breakage unit approximately equal to a 1% chance of a breakage between two markers. One cR is very roughly equivalent to 20 kilobases.

Table 14 indicates the ΔCt values for results of epicenter mapping relative to DNA38113-1230, indicating the relative amplification in the region more immediate to the actual location of DNA38113-1230 along Chromosome 19.

Table 11

Framework Markers Along Chromosome 19

Map Position on Chromosome 19	Stanford Human Genome Center Marker Name
S12	AFMa107xc9
S50	SHGC-31335
S105	SHGC-34102
S155	SHGC-16175

Table 12

Epicenter Markers Along Chromosome 19 used for DNA38113-1230

Map Position on Chromosome 19	Stanford Human Genome Center Marker Name	Distance to next Marker (cR ¹)
S42	WI-7289	5
S43	SHGC-32638	28
S44	SHGC-11753 ²	7
DNA38113-1230	-	-
S45	SHGC-14810	37
S46	AFM214YF6	15
S48	SHGC-36583	-

Table 13
Amplification of framework markers relative to DNA38113-1230 (ΔCt)

Tumor	Framework Markers				
	S12	DNA38113-1230	S50	S105	S155
5	LT1	0.16	-0.15	0.06	-0.42
	LT1a	0.05	0.57	-0.27	0.17
	LT2	0.48	0.57	0.41	0.52
	LT3	0.27	0.77	0.83	0.11
	LT4	0.48	0.08	0.67	0.20
	LT6	0.72	0.33	0.74	0.32
	LT7	0.82	0.29	0.85	0.95
10	LT9	0.72	-0.19	0.61	0.19
	LT10	0.82	1.45	0.98	0.62
	CT2	0.25	2.94	0.29	0.37
	CT3	-0.17	1.23	-0.10	0.34
	CT8	0.13	1.45	0.57	0.18
	CT10	0.15	1.72	0.51	-0.01
	CT12	0.13	1.60	0.57	0.41
15	CT14	0.40	2.03	0.39	0.45
	CT15	-0.23	0.68	-0.30	-0.06
	CT16	0.38	1.07	0.31	0.24
	CT17	0.25	0.50	0.71	0.32
					0.09
20					

Table 14
Amplification of epicenter markers relative to DNA38113-1230 (ΔCt)

Tumor	S41	S42	S43	S44	DNA38113-1230	S45	S46	S48
LT1	-1.03	-0.25	-0.18	-0.11	-0.31	0.13	0.26	0.29
LT1a	0.14	-0.30	-0.11	-0.01	0.21	-0.44	0.45	-0.30
LT2	0.03	0.06	0.06	0.12	0.14	0.16	0.11	0.65
LT3	-1.08	-0.08	-0.01	0.11	0.43	-0.37	0.33	0.56
LT4	0.66	-0.14	-0.48	-0.79	-0.28	-0.31	0.04	0.09
LT6	-0.88	-0.08	-0.12	-1.00	0.20	-0.43	0.48	0.63
LT7	0.65	-0.19	-0.19	-0.04	0.04	-0.42	0.43	0.57
LT9	0.66	-0.26	-0.01	-0.14	-0.06	-0.31	-16.48	0.16
LT10	1.16	-0.30	-0.11	-0.31	0.13	-0.33	0.34	0.50
LT11	0.46	0.01	-0.04	-0.86	0.67	0.23	0.24	-0.57
LT12	1.39	-0.01	-0.22	-1.33	1.57	-0.25	0.26	0.07
LT13	1.62	-0.03	0.00	-0.08	1.22	-0.08	0.48	0.14
LT15	1.09	0.20	0.47	0.62	2.47	0.38	0.01	0.44
LT16	1.51	0.04	-0.04	0.29	2.23	0.51	0.50	0.90
LT17	2.12	0.23	0.11	0.20	1.02	0.45	0.46	-0.41
LT18	1.80	-0.11	0.07	-0.70	0.9	0.10	0.00	-0.02
LT22	-0.12	0.06	0.41	-0.11	-0.06	0.34	0.03	0.52
CT1	-0.09	0.33	0.11	0.22	1.38	0.09	-0.25	-0.10
CT2	1.76	0.04	0.30	0.65	2.94	0.18	-0.04	0.01
CT3	1.10	-0.31	-0.24	0.16	1.23	-0.64	0.78	-0.17
CT4	1.63	0.22	0.32	-0.72	2.23	-0.04	0.44	0.72
CT5	2.22	0.02	0.21	0.10	2.51	0.02	0.18	0.24
CT6	0.48	0.20	0.22	-0.63	2.29	0.03	0.14	0.97
CT7	0.93	0.20	0.32	0.14	0.95	-0.01	0.20	0.54
CT8	1.15	-0.50	-0.14	0.15	1.45	-0.31	0.54	0.07
CT9	0.82	0.38	0.64	-0.71	1.59	1.04	0.26	0.93
CT10	1.57	-0.41	-0.03	-0.14	1.72	-0.27	0.04	0.10
CT11	1.49	-0.05	0.07	0.01	3.34	0.54	0.28	0.88
CT12	0.89	-0.09	-0.01	-0.62	1.6	-0.07	1.16	0.92

CT14	2.16	0.32	0.37	0.47	2.03	-0.07	1.21	0.44
CT15	0.64	-0.52	-0.21	-0.12	0.68	-0.61	1.01	0.32
CT16	1.75	-0.31	0.28	0.47	1.07	0.04	1.01	-0.29
CT17	0.77	-0.18	0.13	-0.04	0.5	-0.27	0.93	0.31
CT18	0.91	0.05	0.14	0.60	1.08	0.22	-0.59	0.61

5

PRO715 (DNA52722-1229):

PRO715 was also reexamined with both framework and epicenter mapping. Table 15 indicates the chromosomal localizations of the framework markers that were used for the procedure. The framework markers are located approximately every 20 megabases and were used to control aneuploidy. Table 16 indicates the epicenter mapping markers that were used in the procedure. The epicenter markers were located in close proximity to DNA52722-1229 and are used to determine the relative DNA amplification in the immediate vicinity of DNA52722-1229. The distance between individual markers is measured in centirays, which is a radiation breakage unit approximately equal to a 1% chance of a breakage between two markers. One cR is very roughly equivalent to about 20 kilobases. In Table 16, "BAC" means bacterial artificial chromosome. The ends of a BAC clone which contained the gene of interest were sequenced. TaqMan primers and probes were made from this sequence, which are indicated in the table. BAC clones are typically 100 to 150 Kb, so these primers and probes can be used as nearby markers to probe DNA from tumors. In Table 16, the marker SHGC-31370 is the marker found to be the closest to the location on chromosome 17 where DNA52722-1229 maps.

15
16
17
18
19
20
21
22
23
24
25

Table 15
Framework Markers Used Along Chromosome 17 for DNA52722-1229

Map Position on Chromosome 17	Stanford Human Genome Center Marker Name
Q4	SHGC-31242
Q52	SHGC-35988
Q110	AFM200zf4
Q169	SHGC-32689
Q206	SHGC-11717
Q232	SHGC-32338

Table 16
Epicenter Markers Used on Chromosome 17 in Vicinity of DNA52722-1229

Map Position on Chromosome 17	Stanford Human Genome Marker Name	Distance to next Marker (cR)
Q33	SHGC-35547	18 cR to Q34
120F17FOR1	Marker from forward end of BAC sequence	
120F17FOR2	Marker from forward end of BAC sequence	
DNA52722-1229	-----	
120F17REV1	Marker from reverse end of BAC sequence	
120F17REV2	Marker from reverse end of BAC sequence	
Q34	SHGC-31370	

Table 17 indicates the ΔCt values of the above described framework markers along chromosome 17 relative to DNA52722-1229 for selected tumors.

Table 17
Amplification of Framework Markers Relative to DNA52722-1229

Tumor	Framework Marker						
	Q4	Q52	DNA5272 2-1229	Q110	Q169	Q206	Q232
LT1	0.02	-0.50	-0.04	0.05	-0.32	-0.21	-0.34
LT1a	-0.01	-0.34	0.64	0.23	-0.20	-0.25	-0.15
LT2	0.25	0.15	0.19	0.05	-0.16	-0.14	-0.09
LT3	-0.08	-0.20	0.54	0.56	-0.06	0.32	0.05
LT4	-0.32	-0.45	0.31	0.19	-0.06	-0.12	0.04
LT6	-0.21	-0.38	0.31	0.13	-0.08	-0.30	0.01
LT7	-0.66	-1.02	0.02	0.62	-0.20	0.06	0.16
LT9	-0.03	-0.29	0.46	1.20	-1.75	-0.22	-0.13
LT10	-0.16	-0.09	0.58	0.11	0.01	-0.33	-0.45
LT11	-0.14	0.29	1.03	0.04	0.30	0.52	0.17
LT12	-0.25	-0.68	0.72	0.65	0.86	0.97	0.58

LT13	0.20	0.00	1.37	-0.15	-0.04	0.25	-0.01
LT15	0.11	-0.39	1.75	0.00	-0.02	0.43	-0.19
LT16	-0.07	-0.56	1.11	0.22	0.19	0.68	-0.55
LT17	0.41	-0.09	1.14	0.27	0.22	0.73	0.07
LT18	0.14	-0.22	1.04	0.27	0.35	0.48	-0.03
LT22	-0.07	-0.73	0.00	0.13	-0.02	0.41	0.05
CT2	0.12	-0.47	1.29	-0.19	0.32	---	0.18
CT3	0.05	0.17	1.06	-0.41	0.05	---	-0.06
CT8	0.44	0.14	1.08	0.02	-0.04	---	-0.11
CT10	0.35	0.26	1.60	-0.05	0.00	---	-0.02
CT12	-0.15	-0.46	0.52	-0.13	0.02	---	-0.20
CT14	0.26	-0.59	1.05	-0.01	0.68	---	0.48
CT15	0.55	-0.51	1.36	-0.69	0.11	---	-0.16
CT16	0.09	-0.14	1.06	0.00	0.00	---	-0.15
CT17	0.40	-0.16	1.00	-0.47	0.04	---	-0.29

Table 18 indicates the ΔCt values for the indicated epicenter markers, indicating the relative amplification along chromosome 17 in the immediate vicinity of DNA52722-1229.

Table 18
Amplification of Epicenter Markers Relative to DNA52722-1229

Tumor	Epicenter marker						
	Q33	120F17FOR 1	120F17FOR 2	DNA5272 2-1229	120F17REV1	120F17REV2	Q34
LT1	-0.18	0.11	0.00	0.20	-0.08	0.07	-0.36
LT1a	0.32	-0.06	0.00	0.68	-0.09	-0.20	0.32
LT2	0.06	0.14	0.00	0.27	-0.29	0.16	-0.16
LT3	0.08	-2.06	0.00	0.16	-0.84	-0.38	-0.16
LT4	---	---	---	---	---	---	---
LT6	---	---	---	---	---	---	---
LT7	-0.20	-0.51	0.00	0.23	-0.63	-0.37	-0.41
LT9	0.08	-0.17	0.00	0.59	0.02	-0.66	-0.01
LT10	0.09	0.05	0.00	0.59	-0.22	-0.12	0.36
LT11	0.75	0.09	0.00	1.07	0.43	-0.01	0.63

LT12	0.00	-0.45	0.00	0.63	-0.49	-0.82	0.18	
LT13	0.72	-0.02	0.00	1.29	0.04	0.02	0.66	
LT15	0.75	0.11	0.00	1.33	0.15	-0.19	0.90	
LT16	0.34	-0.41	0.00	1.11	-0.39	-0.89	0.15	
LT17	1.06	0.29	0.00	1.13	-0.26	-0.12	0.90	
5	LT18	0.66	0.11	0.00	1.21	-0.28	0.11	0.47
LT19	-0.09	-0.37	0.00	0.12	-0.53	-0.48	-0.53	
CT1	0.50	0.14	0.00	1.22	0.27	0.43	0.72	
CT2	0.69	-0.47	0.00	0.95	-0.72	-0.17	0.77	
10	CT3	0.87	0.08	0	1.19	-0.06	0.74	0.97
CT4	0.45	-0.11	0	1.26	0.43	0.38	0.79	
CT5	0.36	-0.39	0	1.79	-0.48	0.09	0.95	
CT6	0.41	0.08	0	1.71	-0.21	0.57	0.47	
15	CT7	0.40	0.18	0	1.19	0.31	0.40	0.54
CT8	0.48	0.17	0	0.93	0.23	0.47	0.72	
CT10	0.72	0.15	0	1.86	0.81	0.67	0.97	
20	CT11	0.80	-0.09	0	2.29	0.20	0.25	0.85
CT12	0.01	-0.55	0	0.49	-0.43	-0.09	0.11	
CT14	0.22	-0.36	0	1.05	0.63	0.41	0.40	
CT15	1.06	-0.04	0	1.27	0.74	0.98	1.13	
CT16	0.84	0.06	0	1.03	0.26	0.40	0.91	
CT17	0.80	0.04	0	0.95	0.78	1.29	0.90	
CT18	0.34	0.13	0	1.06	0.06	0.34	0.50	

PRO357 (DNA44804-1248):

25 PRO357 was reexamined with selected tumors from the above initial screen with framework mapping. Table 19 indicate the chromosomal mapping of the framework markers that were used in the present example. The framework markers are located approximately every 20 megabases and were used to control aneuploidy.

PRO357 was also examined with epicenter mapping. The markers indicated in Table 20 are located in close proximity (in the genome) to DNA44804-1248 and are used to assess the relative amplification in the immediate 30 vicinity of chromosome 16 wherein DNA44804-1248 is located. The distance between individual markers is measured in centirays (cR), which is a radiation breakage unit approximately equal to a 1% chance of a breakage between the two markers. One cR is very roughly equivalent to 20 kilobases. The marker SHGC-6154 is the marker found to be the closest to the location on chromosome 16 where DNA44804-1248 maps.

Table 19
Framework markers for DNA44804-1248

Map position on chromosome 16	Stanford Human Genome Center Marker Name
P7	SHGC-2835
P55	SHGC-9643
P99	GATA7B02
P154	SHGC-33727
P208	SHGC-13577

Table 20
Epicenter markers for DNA44804-1248 along chromosome 16

Map position on chromosome 16	Stanford Human Genome Center Marker Name	Distance to next Marker (cR)
P1	AFMA139WG1	6
P3	SHGC-32420	170 (gap)
P4	SHGC-14817	40
P5	SHGC-12265	4
P6	SHGC-6154	33
DNA44804-1248	—	—
P7	SHGC-2835	10
P8	SHGC-2850	9
P9	AFM297yg5	67
P15	CHLC.GATA70B04	—

The ΔCt values of the above described framework markers along chromosome 16 relative to DNA44804-1248

25 is described in Table 21.

Table 21
Amplification of Framework Markers relative to DNA44804-1248 (ΔCt)

Tumor	Framework marker					
	DNA44804-1248	P7	P55	P99	P154	208
LT1	0.25	0.22	-0.17	0.42	0.04	0.43
LT1a	0.90	0.09	-0.10	-0.38	0.29	0.93
LT2	-0.16	0.03	0.19	-0.18	0.18	0.54
LT3	1.15	0.68	0.57	-0.34	-0.03	0.86

LT4	0.19	0.58	0.36	-0.31	0.08	1.14	
LT6	0.28	0.27	-0.11	-0.74	-0.13	0.22	
LT7	0.58	0.63	0.14	0.82	0.09	-0.21	
LT9	0.68	0.63	0.14	0.82	0.09	-0.21	
LT10	1.21	0.52	0.40	-0.39	-0.15	0.77	
5	LT11	1.71	-0.79	1.31	0.73	-0.08	0.90
LT12	1.96	-0.95	0.94	0.00	-0.63	0.18	
LT13	2.32	-0.97	0.94	0.88	-0.04	0.70	
LT15	3.01	-0.54	0.60	0.12	0.14	1.15	
10	LT16	0.67	-0.27	0.57	-0.39	0.08	1.04
LT17	1.64	0.25	1.10	0.28	0.10	0.23	
LT18	0.34	0.09	0.51	0.33	-0.20	-0.09	
LT19	3.03	-0.82	0.63	0.06	0.09	0.55	
15	LT21	1.33	-1.19	1.01	0.11	0.34	0.07

Table 22 indicates the ΔCt values for the results of epicenter mapping relative to DNA44804-1248, indicating the relative amplification in the region more immediate to the actual location of DNA44804-1248 along chromosome 16.

Table 22
Amplification of epicenter markers relative to DNA44804-1248

Tumor	Epicenter marker										
	P1	P3	P4	P5	P6	DNA4480 4-1248	P7	P8	P9	P15	
LT1	0.31	-0.30	0.65	0.05	-0.33	0.16	-0.41	0.20	0.1	0.17	
LT1a	-0.23	-17.67	0.97	-0.65	-1.83	0.56	-0.65	-0.28	-0.27	-0.07	
LT2	0.18	-0.06	0.33	-0.11	-0.38	-0.32	-1.08	-0.31	-0.53	-0.05	
LT3	0.00	0.25	1.07	-0.23	-0.11	0.70	-0.71	-0.12	-0.17	-0.01	
LT4	0.07	-0.25	0.55	-1.15	-1.78	-0.09	-0.82	-0.07	-0.34	-0.07	
LT6	0.24	0.07	0.48	-0.55	-0.34	-0.07	-1.33	-0.41	-0.7	-0.27	
LT7	0.07	-0.07	0.61	-0.19	-0.36	0.29	-0.96	-0.09	-0.26	-0.08	
LT9	0.16	-0.16	0.64	-0.33	-0.14	0.43	-1.01	-0.19	-0.36	-0.21	
LT10	0.47	0.76	-0.30	0.80	-0.09	0.00	-0.85	-0.17	-0.28	-0.07	
30	LT11	0.14	0.14	0.96	-0.02	0.37	1.27	-0.23	0.09	-0.33	-0.07

LT12	-0.12	-0.04	0.84	-1.52	-0.28	1.42	-0.39	-0.38	-1.21	-0.25	
LT13	0.41	-0.02	1.19	-0.34	0.14	1.67	-0.87	-0.22	-0.72	-0.33	
LT15	0.01	0.21	1.30	-0.48	-0.35	2.36	-0.96	-0.36	-0.54	-0.22	
LT16	-0.38	-0.07	0.41	-0.32	-1.22	-0.08	-0.45	-0.25	-0.52	-0.31	
LT17	0.36	0.23	1.39	-1.39	01.37	1.17	-0.39	-0.13	0.52	0.01	
5	LT18	0.17	-0.27	0.04	-0.04	0.18	-0.39	-0.59	-0.25	-0.21	-0.22
LT19	0.11	-0.02	1.27	-0.12	1.27	2.49	-0.30	-0.36	-0.82	-0.40	
LT21	0.28	-0.18	0.85	0.09	0.66	0.85	-0.49	-0.35	-0.27	-0.16	

CONCLUSION:

10 The ΔCt values for the above DNAs in a variety of tumors are reported. A ΔCt of > 1 was typically used as the threshold value for amplification scoring, as this represents a doubling of gene copy. The above data indicates that significant amplification of the tested nucleic acids occurred in primary lung tumors and/or primary colon tumors: Amplification has been confirmed by framework mapping. The framework markers analysis reports the relative amplification of particular chromosomal regions in the indicated tumors, while the epicenter markers analysis gives a more precise reading of the relative amplification in the region immediately in the vicinity of the gene of interest.

15 Amplification has been confirmed by epicenter mapping and the data evidenced significant amplification in primary colon tumors and/or primary lung tumors: Amplification of the closest known epicenter markers does not occur to a greater extent than that of the DNAs tested. This strongly suggests that the DNAs tested are responsible for the amplification of the particular region on the respective chromosome.

20 Because amplification of the DNAs tested occurs in various lung and colon tumors, it is highly probable that these DNAs play a significant role in tumor formation or growth. As a result, antagonists (e.g., antibodies) directed against the proteins encoded by the DNAs tested would be expected to have utility in cancer therapy and as useful diagnostic reagents. The polypeptides encoded by the DNAs tested have utility as diagnostic markers 25 for determining the presence of tumor cells in lung and/or colon tissue samples. The nucleic acid sequences encoding these polypeptides have utility as sources of nucleic acid probes for carrying out the above diagnostic procedures.

EXAMPLE 29: Ability of PRO241 to Stimulate the Release of Proteoglycans from Cartilage (Assay 97)

30 The ability of PRO241 to stimulate the release of proteoglycans from cartilage tissue was tested as follows. A positive result in this assay evidences that the polypeptide is expected to be useful in the therapeutic treatment of various cartilage and/or bone injuries or disorders including, for example, arthritis.

The metacarpophalangeal joint of 4-6 month old pigs was aseptically dissected, and articular cartilage was removed by free hand slicing being careful to avoid the underlying bone. The cartilage was minced and cultured 35 in bulk for 24 hours in a humidified atmosphere of 95% air, 5% CO_2 in serum free (SF) media (DME/F12 1:1) woth 0.1% BSA and 100U/ml penicillin and 100 μ g/ml streptomycin. After washing three times, approximately

100 mg of articular cartilage was aliquoted into micronics tubes and incubated for an additional 24 hours in the above SF media. PRO241 polypeptides were then added at 1% either alone or in combination with 18 ng/ml interleukin-1 α , a known stimulator of proteoglycan release from cartilage tissue. The supernatant was then harvested and assayed for the amount of proteoglycans using the 1,9-dimethyl-methylene blue (DMB) colorimetric assay (Farndale and Buttle, *Biochem. Biophys. Acta* 883:173-177 (1985)). A positive result in this

5 assay indicates that the test polypeptide will find use, for example, in the treatment of sports-related joint problems, articular cartilage defects, osteoarthritis or rheumatoid arthritis.

When PRO241 polypeptides were tested in the above assay, the polypeptides demonstrated a marked ability to stimulate release of proteoglycans from cartilage tissue both basally and after stimulation with interleukin-1 α and at 24 and 72 hours after treatment, thereby indicating that PRO241 polypeptides are useful for stimulating

10 proteoglycan release from cartilage tissue. As such, PRO241 polypeptides are useful for the treatment of sports-related joint problems, articular cartilage defects, osteoarthritis or rheumatoid arthritis.

Example 30: *In Vitro* Antitumor Assay with PRO344 (Assay 161)

The antiproliferative activity of the PRO344 polypeptide was determined in the investigational, disease-oriented *in vitro* anti-cancer drug discovery assay of the National Cancer Institute (NCI), using a sulforhodamine B (SRB) dye binding assay essentially as described by Skehan et al., *J. Natl. Cancer Inst.* 82:1107-1112 (1990). The 60 tumor cell lines employed in this study ("the NCI panel"), as well as conditions for their maintenance and culture *in vitro* have been described by Monks et al., *J. Natl. Cancer Inst.* 83:757-766 (1991). The purpose of this screen is to initially evaluate the cytotoxic and/or cytostatic activity of the test compounds against different

20 types of tumors (Monks et al., *supra*; Boyd, *Cancer: Princ. Pract. Oncol. Update* 3(10):1-12 [1989]).

Cells from approximately 60 human tumor cell lines were harvested with trypsin/EDTA (Gibco), washed once, resuspended in IMEM and their viability was determined. The cell suspensions were added by pipet (100 μ L volume) into separate 96-well microtiter plates. The cell density for the 6-day incubation was less than for the 2-day incubation to prevent overgrowth. Inoculates were allowed a preincubation period of 24 hours at 37°C

25 for stabilization. Dilutions at twice the intended test concentration were added at time zero in 100 μ L aliquots to the microtiter plate wells (1:2 dilution). Test compounds were evaluated at five half-log dilutions (1000 to 100,000-fold). Incubations took place for two days and six days in a 5% CO₂ atmosphere and 100% humidity.

After incubation, the medium was removed and the cells were fixed in 0.1 ml of 10% trichloroacetic acid at 40°C. The plates were rinsed five times with deionized water, dried, stained for 30 minutes with 0.1 ml of

30 0.4% sulforhodamine B dye (Sigma) dissolved in 1% acetic acid, rinsed four times with 1% acetic acid to remove unbound dye, dried, and the stain was extracted for five minutes with 0.1 ml of 10 mM Tris base [tris(hydroxymethyl)aminomethane], pH 10.5. The absorbance (OD) of sulforhodamine B at 492 nm was measured using a computer-interfaced, 96-well microtiter plate reader.

A test sample is considered positive if it shows at least 50% growth inhibitory effect at one or more concentrations. The results are shown in the following Table 23, where the abbreviations are as follows:

35

NSCL = non-small cell lung carcinoma

CNS = central nervous system

Table 23

<u>Test compound</u>	<u>Concentration</u>	<u>Days</u>	<u>Tumor Cell Line Type</u>	<u>Cell Line Designation</u>
PRO344	1.2 nM	2	Leukemia	HL-60 (TB)
PRO344	1.2 nM	6	Renal	UO-31 and CAKI-1
PRO344	14.9 nM	2	Colon	KM-12
5 PRO344	14.9 nM	2	CNS	SF-268
PRO344	14.9 nM	2	Ovarian	OVCAR-4
PRO344	14.9 nM	2	Renal	CAKI-1
PRO344	14.9 nM	2	Breast	MDA-MB-435
PRO344	14.9 nM	6	Leukemia	HL-60 (TB)
10 PRO344	14.9 nM	6	Colon	KM-12
PRO344	14.9 nM	6	CNS	SF-295
PRO344	14.9 nM	6	NSCL	HOP62

The results of these assays demonstrate that PRO344 polypeptides are useful for inhibiting neoplastic growth in a number of different tumor cell types and may be used therapeutically therefor. Antibodies against PRO344 are useful for affinity purification of this useful polypeptide. Nucleic acids encoding PRO344 polypeptides are useful for the recombinant preparation of these polypeptides.

EXAMPLE 31: Inhibition of Vascular Endothelial Growth Factor (VEGF) Stimulated Proliferation of Endothelial Cell Growth (Assay 9)

The ability of various PRO polypeptides to inhibit VEGF stimulated proliferation of endothelial cells was tested. Polypeptides testing positive in this assay are useful for inhibiting endothelial cell growth in mammals where such an effect would be beneficial, e.g., for inhibiting tumor growth.

Specifically, bovine adrenal cortical capillary endothelial cells (ACE) (from primary culture, maximum of 25 12-14 passages) were plated in 96-well plates at 500 cells/well per 100 microliter. Assay media included low glucose DMEM, 10% calf serum, 2 mM glutamine, and 1X penicillin/streptomycin/fungizone. Control wells included the following: (1) no ACE cells added; (2) ACE cells alone; (3) ACE cells plus 5 ng/ml FGF; (4) ACE cells plus 3 ng/ml VEGF; (5) ACE cells plus 3 ng/ml VEGF plus 1 ng/ml TGF-beta; and (6) ACE cells plus 3 ng/ml VEGF plus 5 ng/ml LIF. The test samples, poly-his tagged PRO polypeptides (in 100 microliter volumes), were then added to the wells (at dilutions of 1%, 0.1% and 0.01%, respectively). The cell cultures were incubated for 6-7 days at 37°C/5% CO₂. After the incubation, the media in the wells was aspirated, and the cells were washed 1X with PBS. An acid phosphatase reaction mixture (100 microliter; 0.1M sodium acetate, pH 5.5, 0.1% Triton X-100, 10 mM p-nitrophenyl phosphate) was then added to each well. After a 2 hour incubation at 37°C, the reaction was stopped by addition of 10 microliters 1N NaOH. Optical density (OD) 35 was measured on a microplate reader at 405 nm.

The activity of PRO polypeptides was calculated as the percent inhibition of VEGF (3 ng/ml) stimulated proliferation (as determined by measuring acid phosphatase activity at OD 405 nm) relative to the cells without

stimulation. TGF-beta was employed as an activity reference at 1 ng/ml, since TGF-beta blocks 70-90% of VEGF-stimulated ACE cell proliferation. The results are indicative of the utility of the PRO polypeptides in cancer therapy and specifically in inhibiting tumor angiogenesis. Numerical values (relative inhibition) are determined by calculating the percent inhibition of VEGF stimulated proliferation by the PRO polypeptides relative to cells without stimulation and then dividing that percentage into the percent inhibition obtained by TGF-
5 β at 1 ng/ml which is known to block 70-90% of VEGF stimulated cell proliferation. The results are considered positive if the PRO polypeptide exhibits 30% or greater inhibition of VEGF stimulation of endothelial cell growth (relative inhibition 30% or greater).

The following polypeptide tested positive in this assay: PRO323.

10 EXAMPLE 32: Rod Photoreceptor Cell Survival (Assay 56)

This assay shows that certain polypeptides of the invention act to enhance the survival/proliferation of rod photoreceptor cells and, therefore, are useful for the therapeutic treatment of retinal disorders or injuries including, for example, treating sight loss in mammals due to retinitis pigmentosum, AMD, etc. Sprague Dawley rat pups at 7 day postnatal (mixed population: glia and retinal neuronal cell types) are killed by decapitation following CO₂ anesthesia and the eyes are removed under sterile conditions. The neural retina is dissected away from the pigment epithelium and other ocular tissue and then dissociated into a single cell suspension using 0.25% trypsin in Ca²⁺, Mg²⁺-free PBS. The retinas are incubated at 37°C for 7-10 minutes after which the trypsin is inactivated by adding 1 ml soybean trypsin inhibitor. The cells are plated at 100,000 cells per well in 96 well plates in DMEM/F12 supplemented with N₂. Cells for all experiments are grown at 20 37°C in a water saturated atmosphere of 5% CO₂. After 2-3 days in culture, cells are fixed using 4% paraformaldehyde, and then stained using CellTracker Green CMFDA. Rho 4D2 (ascites or IgG 1:100), a monoclonal antibody directed towards the visual pigment rhodopsin is used to detect rod photoreceptor cells by indirect immunofluorescence. The results are calculated as % survival: total number of calcein - rhodopsin positive cells at 2-3 days in culture, divided by the total number of rhodopsin positive cells at time 2-3 days in culture. The total cells (fluorescent) are quantified at 20x objective magnification using a CCD camera and NIH image software for MacIntosh. Fields in the well are chosen at random.

The following polypeptides tested positive in this assay: PRO243.

EXAMPLE 33: Pericyte c-Fos Induction (Assay 93)

30 This assay shows that certain polypeptides of the invention act to induce the expression of c-fos in pericyte cells and, therefore, are useful not only as diagnostic markers for particular types of pericyte-associated tumors but also for giving rise to antagonists which would be expected to be useful for the therapeutic treatment of pericyte-associated tumors. Specifically, on day 1, pericytes are received from VEC Technologies and all but 5 ml of media is removed from flask. On day 2, the pericytes are trypsinized, washed, spun and then plated onto 35 96 well plates. On day 7, the media is removed and the pericytes are treated with 100 μ l of PRO polypeptide test samples and controls (positive control = DME+5% serum +/- PDGF at 500 ng/ml; negative control = protein 32). Replicates are averaged and SD/CV are determined. Fold increase over Protein 32 (buffer control)

value indicated by chemiluminescence units (RLU) luminometer reading verses frequency is plotted on a histogram Two-fold above Protein 32 value is considered positive for the assay. ASY Matrix: Growth media = low glucose DMEM = 20% FBS + 1X pen strep + 1X fungizone. Assay Media = low glucose DMEM +5% FBS.

The following polypeptides tested positive in this assay: PRO241.

5

EXAMPLE 34: Inhibitory Activity in Mixed Lymphocyte Reaction (MLR) Assay (Assay 67)

This example shows that one or more of the polypeptides of the invention are active as inhibitors of the proliferation of stimulated T-lymphocytes. Compounds which inhibit proliferation of lymphocytes are useful therapeutically where suppression of an immune response is beneficial.

10 The basic protocol for this assay is described in Current Protocols in Immunology, unit 3.12; edited by J E Coligan, A M Kruisbeek, D H Marglies, E M Shevach, W Strober, National Institutes of Health, Published by John Wiley & Sons, Inc.

15 More specifically, in one assay variant, peripheral blood mononuclear cells (PBMC) are isolated from mammalian individuals, for example a human volunteer, by leukopheresis (one donor will supply stimulator PBMCs, the other donor will supply responder PBMCs). If desired, the cells are frozen in fetal bovine serum and DMSO after isolation. Frozen cells may be thawed overnight in assay media (37°C, 5% CO₂) and then washed and resuspended to 3x10⁶ cells/ml of assay media (RPMI; 10% fetal bovine serum, 1% penicillin/streptomycin, 1% glutamine, 1% HEPES, 1% non-essential amino acids, 1% pyruvate). The stimulator PBMCs are prepared by irradiating the cells (about 3000 Rads).

20 The assay is prepared by plating in triplicate wells a mixture of:

100:1 of test sample diluted to 1% or to 0.1%,

50 :1 of irradiated stimulator cells, and

50 :1 of responder PBMC cells.

25 100 microliters of cell culture media or 100 microliter of CD4-IgG is used as the control. The wells are then incubated at 37°C, 5% CO₂ for 4 days. On day 5, each well is pulsed with tritiated thymidine (1.0 mC/well; Amersham). After 6 hours the cells are washed 3 times and then the uptake of the label is evaluated.

30 In another variant of this assay, PBMCs are isolated from the spleens of Balb/c mice and C57B6 mice. The cells are teased from freshly harvested spleens in assay media (RPMI; 10% fetal bovine serum, 1% penicillin/streptomycin, 1% glutamine, 1% HEPES, 1% non-essential amino acids, 1% pyruvate) and the PBMCs are isolated by overlaying these cells over Lympholyte M (Organon Teknika), centrifuging at 2000 rpm for 20 minutes, collecting and washing the mononuclear cell layer in assay media and resuspending the cells to 1x10⁷ cells/ml of assay media. The assay is then conducted as described above.

35 Any decreases below control is considered to be a positive result for an inhibitory compound, with decreases of less than or equal to 80% being preferred. However, any value less than control indicates an inhibitory effect for the test protein.

The following polypeptide tested positive in this assay: PRO361.

EXAMPLE 35: Tissue Expression Distribution

Oligonucleotide probes were constructed from the PRO polypeptide-encoding nucleotide sequences shown in the accompanying figures for use in quantitative PCR amplification reactions. The oligonucleotide probes were chosen so as to give an approximately 200-600 base pair amplified fragment from the 3' end of its associated template in a standard PCR reaction. The oligonucleotide probes were employed in standard quantitative PCR amplification reactions with cDNA libraries isolated from different human adult and/or fetal tissue sources and analyzed by agarose gel electrophoresis so as to obtain a quantitative determination of the level of expression of the PRO polypeptide-encoding nucleic acid in the various tissues tested. Knowledge of the expression pattern or the differential expression of the PRO polypeptide-encoding nucleic acid in various different human tissue types provides a diagnostic marker useful for tissue typing, with or without other tissue-specific markers, for determining the primary tissue source of a metastatic tumor, and the like. The results of these assays are shown in Table 24 below.

Table 24

	<u>Nucleic Acid</u>	<u>Significantly Expressed In</u>	<u>Not Significantly Expressed In</u>
15	DNA34392-1170	liver, kidney, brain, lung	placenta
	DNA39976-1215	brain	lung
	DNA35595-1228		pancreas, brain, kidney, liver
	DNA34436-1238	lung, placenta, brain	testis
	DNA44176-1244	liver	brain, lung
20	DNA44192-1246	kidney	liver
	DNA44804-1248		lung, brain
	DNA41234-1242	lung, liver, kidney	brain
	DNA45410-1250	lung, brain, kidney, liver	
	DNA46777-1253		liver, placenta, brain

25

EXAMPLE 36: Fetal Hemoglobin Induction in an Erythroblastic Cell Line (Assay 107)

This assay is useful for screening PRO polypeptides for the ability to induce the switch from adult hemoglobin to fetal hemoglobin in an erythroblastic cell line. Molecules testing positive in this assay are expected to be useful for therapeutically treating various mammalian hemoglobin-associated disorders such as the various thalassemias. The assay is performed as follows. Erythroblastic cells are plated in standard growth medium at 1000 cells/well in a 96 well format. PRO polypeptides are added to the growth medium at a concentration of 0.2% or 2% and the cells are incubated for 5 days at 37°C. As a positive control, cells are treated with 100 μ M hemin and as a negative control, the cells are untreated. After 5 days, cell lysates are prepared and analyzed for the expression of gamma globin (a fetal marker). A positive in the assay is a gamma globin level at least 2-fold above the negative control.

35 The following polypeptide tested positive in this assay: PRO243.

EXAMPLE 37: *In situ* Hybridization

In situ hybridization is a powerful and versatile technique for the detection and localization of nucleic acid sequences within cell or tissue preparations. It may be useful, for example, to identify sites of gene expression, analyze the tissue distribution of transcription, identify and localize viral infection, follow changes in specific mRNA synthesis and aid in chromosome mapping.

5 *In situ* hybridization was performed following an optimized version of the protocol by Lu and Gillett, Cell Vision 1:169-176 (1994), using PCR-generated ^{33}P -labeled riboprobes. Briefly, formalin-fixed, paraffin-embedded human tissues were sectioned, deparaffinized, deproteinized in proteinase K (20 g/ml) for 15 minutes at 37°C, and further processed for *in situ* hybridization as described by Lu and Gillett, *supra*. A [^{33}P] UTP-labeled antisense riboprobe was generated from a PCR product and hybridized at 55°C overnight. The slides
10 were dipped in Kodak NTB2 nuclear track emulsion and exposed for 4 weeks.

^{33}P -Riboprobe synthesis

6.0 μl (125 mCi) of ^{33}P -UTP (Amersham BF 1002, SA < 2000 Ci/mmol) were speed vac dried. To each tube containing dried ^{33}P -UTP, the following ingredients were added:

2.0 μl 5x transcription buffer
15 1.0 μl DTT (100 mM)
2.0 μl NTP mix (2.5 mM : 10 μl ; each of 10 mM GTP, CTP & ATP + 10 μl H₂O)
1.0 μl UTP (50 μM)
1.0 μl Rnasin
20 1.0 μl DNA template (1 μg)
1.0 μl H₂O
25 1.0 μl RNA polymerase (for PCR products T3 = AS, T7 = S, usually)

The tubes were incubated at 37°C for one hour. 1.0 μl RQ1 DNase were added, followed by incubation at 37°C for 15 minutes. 90 μl TE (10 mM Tris pH 7.6/1mM EDTA pH 8.0) were added, and the mixture was pipetted onto DE81 paper. The remaining solution was loaded in a Microcon-50 ultrafiltration unit, and spun using program 10 (6 minutes). The filtration unit was inverted over a second tube and spun using program 2 (3 minutes). After the final recovery spin, 100 μl TE were added. 1 μl of the final product was pipetted on DE81 paper and counted in 6 ml of Biofluor II.

The probe was run on a TBE/urea gel. 1-3 μl of the probe or 5 μl of RNA Mrk III were added to 3 μl of loading buffer. After heating on a 95°C heat block for three minutes, the gel was immediately placed on ice. 30 The wells of gel were flushed, the sample loaded, and run at 180-250 volts for 45 minutes. The gel was wrapped in saran wrap and exposed to XAR film with an intensifying screen in -70°C freezer one hour to overnight.

^{33}P -Hybridization

A. Pretreatment of frozen sections

35 The slides were removed from the freezer, placed on aluminium trays and thawed at room temperature for 5 minutes. The trays were placed in 55°C incubator for five minutes to reduce condensation. The slides were fixed for 10 minutes in 4% paraformaldehyde on ice in the fume hood, and washed in 0.5 x SSC for 5 minutes,

at room temperature (25 ml 20 x SSC + 975 ml SQ H₂O). After deproteination in 0.5 μ g/ml proteinase K for 10 minutes at 37°C (12.5 μ l of 10 mg/ml stock in 250 ml prewarmed RNase-free RNase buffer), the sections were washed in 0.5 x SSC for 10 minutes at room temperature. The sections were dehydrated in 70%, 95%, 100% ethanol, 2 minutes each.

5 **B. Pretreatment of paraffin-embedded sections**

The slides were deparaffinized, placed in SQ H₂O, and rinsed twice in 2 x SSC at room temperature, for 5 minutes each time. The sections were deproteinated in 20 μ g/ml proteinase K (500 μ l of 10 mg/ml in 250 ml RNase-free RNase buffer; 37°C, 15 minutes) - human embryo, or 8 x proteinase K (100 μ l in 250 ml RNase buffer, 37°C, 30 minutes) - formalin tissues. Subsequent rinsing in 0.5 x SSC and dehydration were performed as described above.

10 **C. Prehybridization**

The slides were laid out in a plastic box lined with Box buffer (4 x SSC, 50% formamide) - saturated filter paper. The tissue was covered with 50 μ l of hybridization buffer (3.75g Dextran Sulfate + 6 ml SQ H₂O), vortexed and heated in the microwave for 2 minutes with the cap loosened. After cooling on ice, 18.75 ml formamide, 3.75 ml 20 x SSC and 9 ml SQ H₂O were added, the tissue was vortexed well, and incubated at 42°C for 1-4 hours.

15 **D. Hybridization**

1.0 \times 10⁶ cpm probe and 1.0 μ l tRNA (50 mg/ml stock) per slide were heated at 95°C for 3 minutes. The slides were cooled on ice, and 48 μ l hybridization buffer were added per slide. After vortexing, 50 μ l ³³P mix were added to 50 μ l prehybridization on slide. The slides were incubated overnight at 55°C.

20 **E. Washes**

Washing was done 2 x 10 minutes with 2xSSC, EDTA at room temperature (400 ml 20 x SSC + 16 ml 0.25M EDTA, V_f=4L), followed by RNaseA treatment at 37°C for 30 minutes (500 μ l of 10 mg/ml in 250 ml RNase buffer = 20 μ g/ml). The slides were washed 2 x 10 minutes with 2 x SSC, EDTA at room temperature. The stringency wash conditions were as follows: 2 hours at 55°C, 0.1 x SSC, EDTA (20 ml 20 x SSC + 16 ml 25 EDTA, V_f=4L).

25 **F. Oligonucleotides**

In situ analysis was performed on a variety of DNA sequences disclosed herein. The oligonucleotides employed for these analyses are as follows.

(1) **DNA44804-1248 (PRO357)**
30 p1 5'-GGATTCTAATACGACTCACTATAGGGCTGCCGCAACCCCTCACTG-3' (SEQ ID NO:111)
p2 5'-CTATGAAATTAAACCCTCACTAAAGGGACCGCAGCTGGGTGACCGTGTA-3' (SEQ ID NO:112)

(2) **DNA52722-1229 (PRO715)**
35 p1 5'-GGATTCTAATACGACTCACTATAGGGCCCGCCACCTCCT-3' (SEQ ID NO:113)
p2 5'-CTATGAAATTAAACCCTCACTAAAGGGACTCGAGACACCACCTGACCCA-3' (SEQ ID NO:114)
p3 5'-GGATTCTAATACGACTCACTATAGGGCCAAGGAAGGCAGGAGACTCT-3' (SEQ ID NO:115)
p4 5'-CTATGAAATTAAACCCTCACTAAAGGGACTAGGGGTGGAATGAAAAG-3' (SEQ ID NO:116)

(3) DNA38113-1230 (PRO327)

p1 5'-GGATTCTAATACGACTCACTATAGGGCCCCCTGAGCTCTCCGTGTA-3' (SEQ ID NO:117)

p2 5'-CTATGAAATTAACCCTCACTAAAGGGAAAGGCTGCCACTGGTCGTAGA-3' (SEQ ID NO:118)

(4) DNA35917-1207 (PRO243)

5 p1 5'-GGATTCTAATACGACTCACTATAGGGCAAGGAGCCGGACCCAGGAGA-3' (SEQ ID NO:119)

p2 5'-CTATGAAATTAACCCTCACTAAAGGGAGGGGGCCCTGGTGCTGAGT-3' (SEQ ID NO:120)

G. Results

In situ analysis was performed on a variety of DNA sequences disclosed herein. The results from these

10 analyses are as follows.

(1) DNA44804-1248 (PRO357)

Low to moderate level expression at sites of bone formation in fetal tissues and in the malignant cells of an osteosarcoma. Possible signal in placenta and cord. All other tissues negative.

Fetal tissues examined (E12-E16 weeks) include: liver, kidney, adrenals, lungs, heart, great vessels, oesophagus, stomach, spleen, gonad, brain, spinal cord and body wall.

Adult human tissues examined: liver, kidney, stomach, spleen, adrenal, pancreas, lung, colonic carcinoma, renal cell carcinoma and osteosarcoma. Acetominophen induced liver injury and hepatic cirrhosis.

Chimp Tissues examined: thyroid, parathyroid, lymph node, nerve, tongue, thymus, adrenal, gastric mucosa and salivary gland.

20 Rhesus Monkey: cerebrum and cerebellum.

(2) DNA52722-1229 (PRO715)

Generalized high signal seen over many tissues - highest signal seen over placenta, osteoblasts, injured renal tubules, injured liver, colorectal liver metastasis and gall bladder.

25 Fetal tissues examined (E12-E16 weeks) include: placenta, umbilical cord, liver, kidney, adrenals, thyroid, lungs, heart, great vessels, oesophagus, stomach, small intestine, spleen, thymus, pancreas, brain, eye, spinal cord, body wall, pelvis and lower limb.

Adult human tissues examined: liver, kidney, adrenal, myocardium, aorta, spleen, lung, skin, chondrosarcoma, eye, stomach, colon, colonic carcinoma, prostate, bladder mucosa and gall bladder.

30 Acetominophen induced liver injury and hepatic cirrhosis.

Rhesus Tissues examined: cerebral cortex (rm), hippocampus (rm)

Chimp Tissues examined: thyroid, parathyroid, lymph node, nerve, tongue, thymus, adrenal, gastric mucosa and salivary gland.

35 (3) DNA38113-1230 (PRO327)

High level of expression observed in developing mouse and human fetal lung. Normal human adult lung, including bronchial epithelium, was negative. Expression in submucosa of human fetal trachea, possibly in

smooth muscle cells. Expression also observed in non-trophoblastic cells of uncertain histogenesis in the human placenta. In the mouse expression was observed in the developing snout and in the developing tongue. All other tissues were negative. Speculated function: Probable role in bronchial development.

Fetal tissues examined (E12-E16 weeks) include: placenta, umbilical cord, liver, kidney, adrenals, thyroid, lungs, heart, great vessels, oesophagus, stomach, small intestine, spleen, thymus, pancreas, brain, eye, spinal

5 cord, body wall, pelvis and lower limb.

Adult tissues examined: liver, kidney, adrenal, myocardium, aorta, spleen, lymph node, pancreas, lung, skin, cerebral cortex (rm), hippocampus (rm), cerebellum (rm), penis, eye, bladder, stomach, gastric carcinoma, colon, colonic carcinoma, thyroid (chimp), parathyroid (chimp) ovary (chimp) and chondrosarcoma.

10 (4) DNA35917-1207 (PRO243)

Cornelia de Lange syndrome (CdLS) is a congenital syndrome. That means it is present from birth. CdLS is a disorder that causes a delay in physical, intellectual, and language development. The vast majority of children with CdLS are mentally retarded, with the degree of mental retardation ranging from mild to severe. Reported IQ's from 30 to 85. The average IQ is 53. The head and facial features include small head size, thin eyebrows which often meet at the midline, long eyelashes, short upturned nose, thin downturned lips, lowset ears and high arched palate or cleft palate. Other characteristics may include language delay, even in the most mildly affected, delayed growth and small stature, low pitched cry, small hands and feet, incurved fifth fingers, simian creases, and excessive body hair. Diagnosis depends on the presence of a combination of these characteristics. Many of these characteristics appear in varying degrees. In some cases these characteristics may not be present or be so mild that they will be recognized only when observed by a trained geneticist or other person familiar with the syndrome. Although much is known about CdLS, recent reports suggest that there is much more to be learned.

In this study additional sections of human fetal face, head, limbs and mouse embryos were examined. No expression was seen in any of the mouse tissues. Expression was only seen with the antisense probe.

25 Expression was observed adjacent to developing limb and facial bones in the perosteal mesenchyme. The expression was highly specific and was often adjacent to areas undergoing vascularization. The distribution is consistent with the observed skeletal abnormalities in the Cornelia de Lange syndrome. Expression was also observed in the developing temporal and occipital lobes of the fetal brain, but was not observed elsewhere. In addition, expression was seen in the ganglia of the developing inner ear; the significance of this finding is

30 unclear.

Though these data do not provide functional information, the distribution is consistent with the sites that are known to be affected most severely in this syndrome.

35 Additionally, faint expression was observed at the cleavage line in the developing synovial joint forming between the femoral head and acetabulum (hip joint). If this pattern of expression were observed at sites of joint formation elsewhere, it might explain the facial and limb abnormalities observed in the Cornelia de Lange syndrome.

EXAMPLE 38: Activity of PRO243 mRNA in Xenopus Oocytes

In order to demonstrate that the human chordin clone (DNA35917-1207) encoding PRO243 is functional and acts in a manner predicted by the *Xenopus chordin* and *Drosophila sog* genes, supercoiled plasmid DNA from DNA35917-1207 was prepared by Qiagen and used for injection into *Xenopus laevis* embryos. Micro-injection of *Xenopus chordin* mRNA into ventrovegetal blastomeres induces secondary (twinned) axes (Sasai et al., *Cell* 79:779-790 (1994)) and *Drosophila sog* also induces a secondary axis when ectopically expressed on the ventral side of the *Xenopus* embryo (Holley et al., *Nature* 376:249-253 (1995) and Schmidt et al., *Development* 121:4319-4328 (1995)). The ability of *sog* to function in *Xenopus* oocytes suggests that the processes involved in dorsoventral patterning have been conserved during evolution.

Methods

10 **Manipulation of Xenopus embryos:**

Adult female frogs were boosted with 200 I.U. pregnant mare serum 3 days before use and with 800 I.U. of human chorionic gonadotropin the night before injection. Fresh oocytes were squeezed out from female frogs the next morning and in vitro fertilization of oocytes was performed by mixing oocytes with minced testis from sacrificed male frogs. Developing embryos were maintained and staged according to Nieuwkoop and Faber, 15 Normal Table of *Xenopus laevis*, N.-H. P. Co., ed. (Amsterdam, 1967).

Fertilized eggs were dejellied with 2% cysteine (pH 7.8) for 10 minutes, washed once with distilled water and transferred to 0.1 X MBS with 5% Ficoll. Fertilized eggs were lined on injection trays in 0.1 X MBS with 5% Ficoll. Two-cell stage developing *Xenopus* embryos were injected with 200 pg of pRK5 containing wild type chordin (DNA35917-1207) or 200 pg of pRK5 without an insert as a control. Injected embryos were kept on trays for another 6 hours, after which they were transferred to 0.1 X MBS with 50 mg/ml gentamycin until reaching Nieuwkoop stage 37-38.

20 **Results:**

Injection of human chordin cDNA into single blastomeres resulted in the ventralization of the tadpole. The ventralization of the tadpole is visible in the shortening and kinking of the tail and the expansion of the cement 25 gland. The ability of human chordin to function as a ventralizing agent in *Xenopus* shows that the protein encoded by DNA35917-1207 is functional and influences dorsal-ventral patterning in frogs and suggests that the processes involved in dorsoventral patterning have been conserved during evolution, with mechanisms in common between humans, flies and frogs.

30

Deposit of Material

The following materials have been deposited with the American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209, USA (ATCC):

<u>Material</u>	<u>ATCC Dep. No.</u>	<u>Deposit Date</u>
DNA34392-1170	ATCC 209526	December 10, 1997
DNA35917-1207	ATCC 209508	December 3, 1997
DNA39976-1215	ATCC 209524	December 10, 1997
DNA35595-1228	ATCC 209528	December 10, 1997
DNA38113-1230	ATCC 209530	December 10, 1997

DNA34436-1238	ATCC 209523	December 10, 1997
DNA40592-1242	ATCC 209492	November 21, 1997
DNA44176-1244	ATCC 209532	December 10, 1997
DNA44192-1246	ATCC 209531	December 10, 1997
DNA39518-1247	ATCC 209529	December 10, 1997
5 DNA44804-1248	ATCC 209527	December 10, 1997
DNA52722-1229	ATCC 209570	January 7, 1998
DNA41234-1242	ATCC 209618	February 5, 1998
DNA45410-1250	ATCC 209621	February 5, 1998
DNA46777-1253	ATCC 209619	February 5, 1998

10

These deposit were made under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure and the Regulations thereunder (Budapest Treaty). This assures maintenance of a viable culture of the deposit for 30 years from the date of deposit. The deposits will be made available by ATCC under the terms of the Budapest Treaty, and subject to an agreement 15 between Genentech, Inc. and ATCC, which assures permanent and unrestricted availability of the progeny of the culture of the deposit to the public upon issuance of the pertinent U.S. patent or upon laying open to the public of any U.S. or foreign patent application, whichever comes first, and assures availability of the progeny to one determined by the U.S. Commissioner of Patents and Trademarks to be entitled thereto according to 35 USC § 122 and the Commissioner's rules pursuant thereto (including 37 CFR § 1.14 with particular reference to 20 886 OG 638).

The assignee of the present application has agreed that if a culture of the materials on deposit should die or be lost or destroyed when cultivated under suitable conditions, the materials will be promptly replaced on notification with another of the same. Availability of the deposited material is not to be construed as a license to practice the invention in contravention of the rights granted under the authority of any government in accordance with its patent laws.

The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. The present invention is not to be limited in scope by the construct deposited, since the deposited embodiment is intended as a single illustration of certain aspects of the invention and any constructs that are functionally equivalent are within the scope of this invention. The deposit of material herein does not constitute 25 an admission that the written description herein contained is inadequate to enable the practice of any aspect of the invention, including the best mode thereof, nor is it to be construed as limiting the scope of the claims to the specific illustrations that it represents. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the appended claims.